Article:
Tusso, Sergio, Nieuwenhuis, Bart P S, Sedlazeck, Fritz J et al. (3 more authors) (2019) Ancestral Admixture Is the Main Determinant of Global Biodiversity in Fission Yeast. Molecular Biology and Evolution. 1975–1989. ISSN 0737-4038

https://doi.org/10.1093/molbev/msz126

Submission intended as an article

Ancestral admixture is the main determinant of global biodiversity in fission yeast

Sergio Tusso [1,2,*], Bart P.S. Nieuwenhuis [1], Fritz J. Sedlazeck [3], John W. Davey [4], Daniel C. Jeffares [5,6], Jochen B. W. Wolf[1,2,*]

[1] Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

[2] Science for Life Laboratories and Department of Evolutionary Biology, Norbyvägen 18D, 75236 Uppsala University, 75236 Uppsala, Sweden

[3] Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030

[4] Bioscience Technology Facility, Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom

[5] Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom

[6] York Biomedical Research Institute (YBRI), University of York, Wentworth Way, York YO10 5DD, United Kingdom

[*] Authors to whom correspondence should be addressed

Corresponding authors:

Sergio Tusso: situssog@gmail.com

Jochen B. W. Wolf: j.wolf@biologie.uni-muenchen.de

Mutation and recombination are key evolutionary processes governing phenotypic variation and reproductive isolation. We here demonstrate that biodiversity within all globally known strains of *Schizosaccharomyces pombe* arose through admixture between two divergent ancestral lineages. Initial hybridization was inferred to have occurred ~20-60 sexual outcrossing generations ago consistent with recent, human-induced migration at the onset of intensified transcontinental trade. Species-wide heritable phenotypic variation was explained near-exclusively by strain-specific arrangements of alternating ancestry components with evidence for transgressive segregation. Reproductive compatibility between strains was likewise predicted by the degree of shared ancestry. To assess the genetic determinants of ancestry block distribution across the genome, we characterized the type, frequency and position of structural genomic variation (SV) using nanopore and single-molecule real time sequencing. Despite being associated with double-strand break initiation points, over 800 segregating structural variants exerted overall little influence on the introgression landscape or on reproductive compatibility between strains. In contrast, we found strongly increased statistical linkage between ancestral populations consistent with negative epistatic selection shaping genomic ancestry combinations during the course of hybridization. This study provides a detailed, experimentally tractable example that genomes of natural populations are mosaics reflecting different evolutionary histories. Exploiting genome-wide heterogeneity in the history of ancestral recombination and lineage-specific mutations sheds new light on the population history of *S. pombe* and highlights the importance of hybridization as a creative force in generating biodiversity.

Introduction

Mutation is the ultimate source of biodiversity. In sexually reproducing organisms it is assisted by recombination shuffling mutations of independent genomic backgrounds into millions of novel combinations. This widens the phenotypic space upon which selection can act and thereby accelerates evolutionary change (Muller, 1932; Fisher, 1999; McDonald et al., 2016). This effect is enhanced for heterospecific recombination between genomes of divergent populations (Abbott et al., 2013). Novel combinations of independently accumulated mutations can significantly increase the overall genetic and phenotypic variation, even beyond the phenotypic space of parental lineages (transgressive segregation (Lamichhaney et al., 2017; Nolte and Sheets, 2005)). Yet, if mutations of the parental genomes are not compatible to produce viable and fertile offspring, hybridization is a dead end. Phenotypic variation then remains within the confines of genetic variation of each reproductively isolated, parental lineage.

It is increasingly recognised that hybridization is commonplace in nature, and constitutes an important driver of diversification (Abbott et al., 2013; Mallet, 2005). Ancestry components of hybrid genomes

can range from clear dominance of alleles from the more abundant species (Dowling et al., 1989; Taylor and Hebert, 1993), over a range of admixture proportions (Lamichhaney et al., 2017; Runemark et al., 2018) to the transfer of single adaptive loci (The Heliconius Genome Consortium et al., 2012). The final genomic composition is determined by a complex interplay of demographic processes, heterogeneity in recombination (e.g. induced by genomic rearrangements) (Wellenreuther and Bernatchez, 2018) and selection (Sankararaman et al., 2014; Schumer et al., 2016). Progress in sequencing technology, now allows characterisation of patterns of admixture and the illumination of underlying processes (Payseur and Rieseberg, 2016). Yet, research has largely focused on animals (Turner and Harr, 2014; Vijay et al., 2016; Meier et al., 2017; Jay et al., 2018) and plants (Twyford et al., 2015) and relatively little attention has been paid to natural populations of sexually reproducing micro-organisms (Leducq et al., 2016; Stukenbrock, 2016; Peter et al., 2018; Steenkamp et al., 2018).

The fission yeast *Schizosaccharomyes pombe* is an archiascomycete haploid unicellular fungus with a facultative sexual mode of reproduction. Despite of its outstanding importance as a model system in cellular biology (Hoffman et al., 2015) and the existence of global sample collections, essentially all research has been limited to a single isogenic strain isolated by Leupold in 1949 (Leupold 972; JB22 in this study). Very little is known about the ecology, origin, and evolutionary history of the species (Jeffares, 2018). Global population structure has been described as shallow with no apparent geographic stratification (Jeffares et al., 2015). Genetic diversity ($\pi = 3 \cdot 10^{-3}$ substitutions/site) appears to be strongly influenced by genome-wide purifying selection with the possible exception of region-specific balancing selection (Fawcett et al., 2014; Jeffares et al., 2015). Despite the overall low genetic diversity, *S. pombe* shows abundant additive genetic variation in a variety of phenotypic traits including growth, stress responses, cell morphology, and cellular biochemistry (Jeffares et al., 2015). The apparent worldwide lack of genetic structure in this species appears inconsistent with the large phenotypic variation between strains and with evidence for post-zygotic reproductive isolation between inter-strain crosses, ranging from 1% to 90 % of spore viability (Kondrat'eva and Naumov, 2001; Teresa Avelar et al., 2013; Zanders et al., 2014; Jeffares et al., 2015; Naumov et al., 2015; Marsellach, 2017).

In this study, we integrate whole-genome sequencing data from three different technologies - sequencing-by-synthesis (Illumina technology data accessed from (Jeffares et al., 2015)), single-molecule real-time sequencing (Pacific BioSciences technology, this study) and nanopore sequencing (Oxford Nanopore technology, this study) - sourced from a mostly human-associated, global sample collection to elucidate the evolutionary history of the *S. pombe* complex. Using population genetic analyses based on single nucleotide polymorphism (SNP) we show that global genetic variation and heritable phenotype variation of *S. pombe* results from recent hybridization of two ancient lineages. 25 *de novo* assemblies from 17 divergent strains further allowed us to quantify segregating structural

variation including insertions, deletions, inversion and translocations. In light of these findings, we retrace the global population history of the species, and discuss the relative importance of genome-wide ancestry and structural mutations in explaining phenotypic variation and reproductive isolation.

Results

Global genetic variation in *S. pombe* is characterized by ancient admixture

Genetic variation of the global *S. pombe* collection comprises 172,935 SNPs segregating in 161 strains. Considering SNPs independently, individuals can be sub-structured into 57 clades that differ by more than 1900 variants, but are near-clonal within clades (Jeffares et al., 2015). To examine population ancestry further, we divided the genome into 1925 overlapping windows containing 200 SNPs each and selected one representative from each clade (57 samples in total). Principle component analysis conducted on each orthologous window showed a highly consistent pattern along the genome (**Figure 1a, Supplementary Figure 1**): i) the major axis of variation (PC1) split all samples into two clear discrete groups explaining 60% ± 13% of genetic variance (**Figure 1b**). ii) All samples fell into either extreme of the normalized distribution of PC1 scores ($PC1 \in [0; 0.25] \cup [0.75; 1]$) (**Supplementary Figures 2 & 3**) with the only exception of strains with inferred changes in ploidy level (**Methods, Supplementary Figure 4**). iii) PC2 explained 13% ± 6% of variation and consistently attributed higher variation to one of the two groups. This strong signal of genomic windows separating into two discrete groups suggested that the genomic diversity in this collection was derived from two distinct ancestral populations. However, iv) group membership of strains changes among windows moving along the genome, reflecting recombination between these two well defined groups. Forwards population simulations followed by principle component analysis with SNP windows showed that this signal was unlikely to be an artefact (see **Methods).** This last point highlights the importance of considering haplotype structure and explains the lack of observed population structure when disregarding non-independence of SNPs (Jeffares et al., 2015).

The strong signal from the PCA systematically differentiating groups along the genome was likewise reflected in population genetic summary statistics including Watterson's theta (θ), pairwise nucleotide diversity (π), and Tajima's D (**Figure 1d and 2**). Significant differences in these statistics (Kendall's τ p-value $\leq 2.2 \cdot 10^{-16}$) were also present in mitochondrial genetic variation (**Figure 1a**) and allowed polarising the two groups across windows into a 'low-diversity' group (red) and a 'high-diversity group (blue) (**Figure 1a**, **Supplementary Figure 5)**. Genetic divergence between groups (D_{xy}) was 15 and 3 times higher than mean genetic diversity (π) within each group, respectively, and thus supports a period of independent evolution. Painting genomic windows by group membership revealed blocks of ancestry distributed in sample specific patterns along the genome (**Figure 1c, Supplementary Figure 6**). The sample corresponding to the reference genome isolated originally from Europe (Leupold's

147 972; JB22) consisted almost exclusively of 'red' ancestry (>96% red), whereas other samples were
148 characterized near-exclusively by 'blue' ancestry (>96% blue). The sample considered to be a different
149 species from Asia, *S. kambucha* (JB1180 (Singh and Klar, 2002)) had a large proportion of 'blue'
150 windows (>70% blue). Hereafter, we refer to the 'red' and 'blue' clade as *Sp* and *Sk*, for *S. pombe* and *S.*
151 *kambucha* respectively.
152
153 Next, we explored different historical processes that may underlie the deep divergence between these
154 clades and their distribution along the genome. Using individual based forward simulations we
155 contrasted a historical scenario of divergence in allopatry followed by extensive, recent hybridization
156 with a scenario of low, but constant levels of gene flow during divergence. The empirical level of
157 divergence between ancestral groups in combination with the observed distribution of recombinant
158 block size along the genome could not be recovered under a demographic model of divergence with
159 constant gene flow. The simulations much rather provide support for a scenario of divergence in
160 isolation followed by a recent pulse of gene flow recovering both ancestral divergence and the
161 observed introgression landscape (see Material and Methods, **Supplementary Figures 7 and 8**).
162
163 The distribution of ancestry components was highly heterogeneous across the chromosome (**Figure**
164 **2a**). Most strains showed an excess of *Sp* ancestry in parts of chromosome I, whereas several regions
165 of chromosome III had an excess of *Sk* ancestry. Grouping samples by the pattern of genomic ancestry
166 across the genome revealed eight discrete consistent clusters (**Figure 1c**). Consistent with independent
167 and/or recent segregation of ancestral groups, cluster membership for several samples differed
168 between chromosomes (**Figure 1c**) and genome components (**Supplementary Figures 9 & 10**). This
169 is also reflected by low support in the relationship between the eight clusters. Failing to incorporate
170 this genome-wide variation of admixture proportions can mimic signatures of selection. For example,
171 equal ancestry contributions for a certain genomic region will yield high positive values of both
172 Tajima's D (**Supplementary Figure 11**) and π and may be mistaken as evidence for balancing
173 selection. Strong skew in ancestry proportions reduces both statistics to values of the prevailing
174 ancestry and may appear as evidence for selective sweeps (**Figure 2b**). Taking ancestry into account,
175 however, there was no clear signature of selection in either *Sp* or *Sk* genetic variation that could
176 account for heterogeneity in the genetic composition of hybrids (**Supplementary Figure 11)**.
177 Signatures of selection identified previously (cf. Fawcett et al., 2014) are likely artefacts due to
178 skewed ancestry proportions rather than events of positive or balancing selection in the ancestral
179 populations.
180
181 Overall, our results provide strong evidence for the presence of at least two divergent ancestral
182 populations: one genetically diverse group (*Sk* clade) and a less diverse group (*Sp* clade). We found a
183 large range of ancestral admixture proportions between these two clades broadly clustering samples

into 8 weakly supported groups. These resemble clusters of strains previously identified by *Structure* and *fineStructure* without explicit modelling of ancestral admixture (Jeffares et al., 2015). Neglecting ancestry, Jeffares et al. (2015) argued that the shallow population structure likely results from extensive gene flow between clusters. Yet, considering the genome-wide distribution of discrete *Sk* and *Sp* ancestry, and lack of geographical structure, suggests that the 8 clusters are derived from one or a few centres of ancient admixture (hybridization) without having to invoke subsequent or recent gene flow between them.

Age of ancestral lineages and timing of hybridization

To shed further light on the population history, we estimated the age of the parental lineages and the timing of initial hybridization. Calibrating mitochondrial divergence by known collection dates over the last 100 years, Jeffares et al. (2015) estimated that the time to the most recent common ancestor for all samples was around 2,300 years ago. Current overrepresentation of near-pure *Sp* and *Sk* in Europe or Africa / Asia, respectively, is consistent with an independent history of the parental lineages on different continents for the most part of the last millennia (**Supplementary Figure 10**). Yet, the variety of admixed genomes bears testimony to the fact that isolation has been disrupted by heterospecific gene flow. Using a theoretical model assuming secondary contact with subsequent hybridization (Janzen et al., 2018) as supported by the data, we estimated that hybridization occurred within the last around 20 to 60 sexual outcrossing generations depending on window size (**Figure 3, Supplementary Figures 12 & 13**). Considering intermittent generations of asexual reproduction, high rates of haploid selfing and dormancy of spores (Farlow et al., 2015; Jeffares, 2018) it is difficult to obtain a reliable estimate of time in years. Yet, the range of estimates of hybridization timing is consistent with hybridization induced by the onset of regular trans-continental human trade between Europe with Africa and Asia (~14th century) and with the Americas (~16th century), with fission yeast as a human commensal (Jeffares, 2018). This fits with the observation that all current samples from the Americas were hybrids, while samples with the purest ancestry stem from Europe, Africa and Asia. Moreover, negative genome-wide Tajima's D estimates for both ancestral clades (mean ± SD for Sp: -0.8 ± 0.9 and Sk: -0.7 ± 0.6) signal a period of recent expansion.

Heritable phenotypic variation and reproductive isolation are governed by ancestry components

Hybridization can lead to rapid evolution due to selection acting on the genetic and phenotypic variation emerging after admixture (Muller, 1932; Fisher, 1999; McDonald et al., 2016). We assessed the consequences of hybridization on phenotypic variation making use of a large data set including 228 quantitative traits collected from the strains under consideration here (Jeffares et al., 2015). Contrary to genetic clustering of hybrid genomes (**Figure 1c**), samples with similar ancestry proportions did not group in phenotypic space described by the first two PC-dimensions capturing 31% of the total variance across traits (**Figure 4a**). Moreover, phenotypic variation of hybrids

exceeded variation of pure strains (>0.9 ancestry for *Sp* or *Sk*). This was supported by trait specific analyses. We divided samples into three discrete groups: pure *Sp*, pure *Sk* and hybrids with a large range of *Sp* admixture proportions (0.1-0.9). 63 traits showed significant difference among groups (**Figure 4b, Supplementary Figure 14**). In the vast majority of cases (50 traits), hybrid phenotypes were indistinguishable from one of the parents, but differed from the other, suggesting dominance of one ancestral background, consistent with some ecological separation of the backgrounds. In seven traits, hybrid phenotypes were intermediate differing from both parents, consistent with an additive or polygenic contribution of both ancestral backgrounds. For six traits, hybrids exceeded phenotypic values of both parents providing evidence for transgressive segregation. In all cases, the number of significantly differentiated traits was found to be higher than under the null model (mean number of significant traits after 10000 permutations: dominant Sk 4 +/- 2, dominant Sk 4 +/- 2, transgressive 0 +/- 0.3, intermediate 0 +/- 0.1; **Supplementary Figure 15**). Jeffares et al. (2017) showed that for each trait the total proportion of phenotypic variance explained by the additive genetic variance component (used as an estimated of the narrow-sense heritability) ranged from 0 to around 90%. We found that across all 228 traits, considering *Sp* and *Sk* ancestry components across the 1,925 genomic windows explained an equivalent amount of phenotypic variance as all 172,935 SNPs segregating across all samples, being both highly correlated (**Figure 4c, 4d;** $r = 0.82$, p-value $\leq 2.2 \cdot 10^{-16}$). Combinations of ancestral genetic variation appear to be the main determinants of heritable phenotypic variation with only little contribution from single-nucleotide mutations arising after admixture. In turn, this supports that the formation of hybrids is recent (see above), and few (adaptive) mutations have occurred after it.

Ancestry also explained most of the variation in postzygotic reproductive isolation between strains. Previous work revealed a negative correlation between spore viability and genome-wide SNP divergence between strains (Jeffares et al., 2015). The degree of similarity in genome-wide ancestry had the same effect: the more dissimilar two strains were in their ancestry, the lower the viability of the resulting spores (**Figure 4e**; Kendall correlation coefficients, $\tau = -0.30$, T= 259, p-value = $6.66 \cdot 10^{-3}$). This finding is consistent with reproductive isolation being governed by many, genome-wide incompatibilities between the *Sp* and *Sk* clade. Yet, in a number of cases spore survival was strongly reduced in strain combinations with near-identical ancestry. In these cases, reproductive isolation may be caused by few large effect mutations, including structural genomic changes that arose after hybridization.

Structural mutations do not determine the genome-wide distribution of ancestry blocks

Structural genomic changes (structural variants or SVs hereafter) are candidates for large-effect mutations governing phenotypic variation (Küpper et al., 2016; Jeffares et al., 2017), reproductive isolation (Hoffmann and Rieseberg, 2008; Teresa Avelar et al., 2013) and heterospecific recombination (Ortiz-Barrientos et al., 2016). They may thus importantly contribute to shaping

heterogeneity in the distribution of ancestry blocks observed along the genome (Jay et al., 2018; Poelstra et al., 2014) (**Figure 2b)**. However, inference of SVs in natural strains of fission yeast has been primary based on short-read sequencing (Jeffares et al., 2017). SV calls from short-read sequencing data are known to differ strongly by bioinformatic pipeline, are prone to false positive inference and are limited in their ability to infer long-range SVs, in particular in repetitive regions of the genome (Jeffares et al., 2017).

To obtain a reliable and comprehensive account of SVs segregating across strains and test for a possible association of SVs with the skewed ancestry in the genome, we generated chromosome-level *de novo* genome assemblies for 17 of the most divergent samples using single-molecule real time sequencing (mean sequence coverage 105x; **Supplementary Table 7**). For the purpose of methodological comparison, we also generated *de novo* assemblies for a subset of 8 strains (including the reference Leupold's 972) based on nanopore sequencing (mean sequence coverage: 140x). SVs were called using a mixed approach combining alignment of *de novo* genomes and mapping of individual reads to the reference genome (Wood et al., 2002). Both approaches and technologies yielded highly comparable results (**Methods, Supplementary Figure 16-19 and Supplementary Table 8).**

After quality filtering, we retained a total of 832 variant calls including 563 insertions or deletions (indels), 118 inversions, 110 translocations and 41 duplications. The 17 strains we examined with long reads could be classified into six main karyotype arrangements (**Figure 5a**). The previously reported list of SVs of the same strains using short reads consisted of only 52 SVs (Jeffares et al., 2017) of which only 8 were found to overlap with the 832 calls from long-read data. The vast majority of SVs were smaller than 10 kb (**Figure 5b**). The size distribution was dominated by elements of 6 kb and 0.5 kb in length corresponding to known transposable elements (TEs) and their flanking long terminal regions (LTRs), respectively (Kelly and Levin, 2005). Only a small number of SVs corresponded to large-scale rearrangements (50 kb - 2.2 Mb) including translocations between chromosomes (**Figure 5a**). A subset of these have been characterized previously as large-effect modifiers of recombination promoting reproductive isolation (Brown et al., 2011; Teresa Avelar et al., 2013; Jeffares et al., 2017).

Contrary to previous SV classification based on short reads (Jeffares et al., 2017), SV density was not consistently increased in repetitive sequences such as centromeric and telomeric regions illustrating the difficulty of short-read data in resolving SVs in repetitive regions (**Figure 5c**). Instead, we found that the frequency of SVs was significantly elevated in close proximity to developmentally programmed DNA double-strand breaks (DSB) associated with recombination initiation (Fowler et al., 2014). The proportion of SVs observed within [0, 0.5) kb and [0.5, 1) kb of DSB was increased by 46% (p-value$<1\cdot10^{-4}$) and 67% (p-value$<1\cdot10^{-4}$) relative to random expectations. On the contrary,

regions more distant than 10 kb from DSB were relatively depleted of SVs (**Supplementary Figure 20**).

Next, we imputed the ancestry of SV alleles from SNPs surrounding SV break points. We calculated allele frequencies for SVs in both ancestral clades and constructed a folded two-dimensional site frequency spectrum (**Figure 5d**). The majority of variants (66 %) segregated at frequencies below 0.3 in both ancestral genetic backgrounds. Very few SVs were differentiated between ancestral populations (3 % of variants with frequency higher than 0.9 in one population and below 0.1 in the other). This pattern contrasted with the reference spectrum derived from SNPs where the proportion of low frequency variants was similar at 60 %, but genetic differentiation between populations was substantially higher (21 % of SNP variants with frequency higher than 0.9 in one population and below 0.1 in the other). The difference was most pronounced for large SVs (larger than 10 kb) and TEs, for which we estimated allele frequencies for all 57 strains by means of PCR and short-read data, respectively. For TE's, 98 % of the total 1048 LTR variants segregated at frequencies below 0.3 in both ancestral populations without a single variant differentiating ancestral populations (**Figure 5d**). Large SVs likewise segregated at low frequencies, being present at most in two strains out of 57. This included the translocation reported for *S. kambucha* between chromosome II and III (Zanders et al., 2014), which we found to be specific for that strain. Only the large inversion on chromosome I segregated at higher frequency being present in five strains out of 57, of which three were of pure *Sp* ancestry including the reference strain (**Supplementary Table 10**). Additionally, SVs segregating at high frequency (> 0.7) did not cluster in genomic regions with steep transitions in ancestry between *Sp* and *Sk* ancestry (large changes in ancestral frequency in **Figure 2a**. p-value > 0.1; **Supplementary Figure 21**).

In summary, long-read sequencing provided a detailed account of species-wide diversity in structural genetic variation including over 800 high-quality variants ranging from small indels to large-scale inter-chromosomal rearrangements. SV calls showed substantial overlap among technologies (Pacific Biosciences, Nanopore) and approaches (de novo assembly vs. mapping), but less than 1 % of this variation was inferred from short-read data. This finding admonishes to caution when interpreting SV calls from short read data which are moreover sensitive to genotyping methods. In contrast to genome-wide SNPs, SVs segregated near-exclusively at low frequencies and were rarely differentiated by ancestral origin. This is consistent with strong diversity-reducing purifying selection relative to SNPs. The fact that SVs, including large-scale rearrangements with known effects on recombination and reproductive isolation (Brown et al., 2011; Teresa Avelar et al., 2013; Zanders et al., 2014), were often unique to single strains precludes a role of SVs in shaping patterns of ancestral heterospecific recombination. Moreover, while being concentrated in proximity to double-strand breaks, possibly due to improper repair upon recombination (Currall et al., 2013), SVs were not significantly associated

with steep transitions in ancestry blocks. Summarizing the evidence, SVs appear to have had limited influence in shaping genome-wide patterns of ancestral admixture. While they may reduce fitness of hybrids between specific strains, SV cannot explain the prevalence of reproductive isolation as a function on ancestral similarity (**Figure 4e**).

Negative epistasis and the distribution of ancestral blocks

Alternatively, heterogeneity in the distribution and frequency of ancestry along the genome may result from negative epistatic interactions of incompatible genetic backgrounds (Schumer et al., 2016). An excess of homospecific combinations of physically distant loci can serve as an indication of epistatic selection against genetic incompatibilities which can be segregating at appreciable frequencies even within species (Corbett-Detig et al., 2013). We tested this hypothesis by measuring ancestry disequilibrium (AD) between all possible pairs of genomic windows within a chromosome. Specifically, we quantified linkage disequilibrium (LD) between windows dominated by alleles from the same ancestral group (> 0.7) *Sp*-*Sp* or *Sk*-*Sk* (reflecting positive AD) and contrasted it to the degree of linkage disequilibrium arising between heterospecific allele combinations *Sp*-*Sk* (negative AD) (**Supplementary Figure 22**). LD differed significantly between these two cases (**Figure 6**). While negative AD decreased rapidly with genetic distance ($R^2 < 0.2$ after 66, 19 and 21 kb respectively for each chromosome) positive LD was higher in magnitude and extended over larger distances ($R^2 < 0.2$ after 1.02, 0.54, and 0.18 Mb respectively for each chromosome in *Sk*-*Sk* comparisons and 1.59, 1.12, and 0.32 Mb for *Sp*-*Sp* comparisons). This relationship remained significant after controlling for the potential effect of secondary population structure which can likewise increase apparent AD mimicking signatures of selection on epistatic genetic variation. Choosing only a single representative from each of the hybrid clusters (**Figure 1c**) LD remained larger for comparisons of genomic regions with the same ancestry. These results are overall consistent with a contribution of epistatic selection during the course of hybridization in shaping the ancestry composition of admixed genomes.

Discussion

This study adds to the increasing evidence that hybridization plays an important role as a rapid, 'creative' evolutionary force in natural populations (Seehausen, 2004; Mallet, 2007; Soltis and Soltis, 2009; Abbott et al., 2013; Schumer et al., 2014; Abbott et al., 2016; Pennisi, 2016; Nieto Feliner et al., 2017). Recent heterospecific recombination between two ancestral *S. pombe* populations shuffled genetic variation of genomes that diverged since classical antiquity about 2,300 years ago. With the necessary caution the timing of hybridization can be inferred to coincide with the onset of intensified trans-continental human trade, suggesting an anthropogenic contribution. Several samples showed similar distribution of ancestral blocks along the genome suggesting comparable evolutionary histories, and allowing the identification of 8 discrete clusters. These clusters, in general showed weak geographical grouping, initially interpreted as evidence for reduced population structure with large

recent world-wide gene flow (Jeffares et al., 2015). In contrast, the world-wide distribution of the two ancestral linages suggests rapid and recent global dispersion after hybridization followed by local differentiation. This study thus highlights the importance of taking genomic non-independence into account. Allowing for the fact that genomes are mosaics reflecting different evolutionary histories can fundamentally alter inference on a species' evolutionary history.

Moreover, conceptualizing genetic variation as a function of ancestry blocks alternating along the genome changes the view on adaptation. Admixture is significantly faster than evolutionary change solely driven by mutation. Accordingly, phenotypic variation was near-exclusively explained by ancestry components with only little contribution from novel mutations. Importantly, admixture not only filled the phenotypic space between parental lineages, but also promoted transgressive segregation in several hybrids. This range of phenotypic outcomes opens the opportunity for hybrids to enter novel ecological niches (Nolte and Sheets, 2005; Pfennig et al., 2016) and track rapid environmental changes (Eroukhmanoff et al., 2013).

Structural mutations have been described as prime candidates for rapid large-effect changes with implications on phenotypic variation, recombination and reproductive isolation (Faria and Navarro, 2010; Ortiz-Barrientos et al., 2016; Wellenreuther and Bernatchez, 2018). This study contributes to this debate providing a detailed account of over 800 high-quality structural variants identified across 17 chromosome level *de nov*o genomes sampled from the most divergent strains within the species. On the whole, SVs had little effect in our analysis. SVs segregated at low frequencies in both ancestral populations and, contrary to what has been suggested for specific genomic regions in other systems (Jay et al., 2018), they did not account for genome-wide heterogeneity in introgression among strains. Moreover, reproductive isolation was overall best predicted by the degree of shared ancestry with little contribution from SVs. Few crosses, however, showed strong reproductive isolation despite a high degree of shared ancestry (outliers in lower left corner of **Figure 4e**). In these cases, combining SVs from different strains into hybrid genomic backgrounds may have a significant impact. This is consistent with the observation that large, artificially generated rearrangements affect fitness (Teresa Avelar et al., 2013; Nieuwenhuis et al., 2018) and may promote reproductive isolation between specific *S. pombe* strains (Teresa Avelar et al., 2013; Zanders et al., 2014). Thus, reproductive isolation may arise by a combination of factors: negative epistasis between many loci with small to moderate effect differing in ancestry, and for specific strain combinations, single major effect (structural) mutations such as selfish elements or meiotic drivers (Zanders et al., 2014; Hu et al., 2017; Nuckolls et al., 2017). Functional work is needed to isolate these genetic elements.

406 **Material and Methods**

407

408 **Strains**

409 This study is based on a global collection of *S. pombe* consisting of 161 world-wide distributed strains
410 (see **Supplementary Table 1**) described in Jeffares *et al.* (2015).

411

412 **Inferring ancestry components**

413 To characterize genetic variation across all strains, we made use of publically available data in variant
414 call format (VCF) derived for all strains from Illumina sequencing with an average coverage of around
415 80x (Jeffares et al., 2015). The VCF file consists of 172,935 SNPs obtained after read mapping to the
416 *S. pombe* 972 h^- reference genome (ASM294v264) (Wood et al., 2003) and quality filtering (see
417 **Supplementary Table 1** for additional information). We used a custom script in R 3.4.3 (Team,
418 2014) with the packages *gdsfmt* 1.14.1 and *SNPRelate* 1.12.2 (Zheng et al., 2017, 2012), to divide the
419 VCF file into genomic windows of 200 SNPs with overlap of 100 SNPs. This resulted in 1925
420 genomic windows of 1 - 89 kb in length (mean 13 kb). For each window, we performed principal
421 component analyses (PCA) using *SNPRelate* 1.12.2 (Zheng et al., 2017, 2012) (example in **Figure 1a**
422 and **Supplementary Figure 1**). The proportion of variance explained by the major axis of variation
423 (PC1) was consistently high and allowed separating strains into two genetic groups/clusters, *Sp* and *Sk*
424 (see main text, **Figure 1b**). Using individual based forward simulations with *SLiM 3.2.1* (Haller and
425 Messer, 2019) we found that this pattern cannot be obtained in the absence of population structure
426 even for smaller window sizes including 5, 10, 20, 50 or 100 SNPs (**Supplementary Figures 23**).

427

428 We calculated population genetic parameters within clusters including pairwise nucleotide diversity
429 (π) (Nei and Li, 1979), Watterson theta (θ_w) (Watterson, 1975), and Tajima's *D* (Tajima, 1989), as
430 well as the average number of pairwise differences between clusters (D_{xy}) (Nei and Li, 1979) using
431 custom scripts. Statistical significance of the difference in nucleotide diversity (π) between ancestral
432 clades was inferred using Kendall's τ as test statistic. Since values of adjacent windows are
433 statistically non-independent due to linkage, we randomly subsampled 200 windows along the genome
434 with replacement. This was repeated a total of 10 times for each test statistic, and we report the
435 maximum p-value. Given the consistent difference between clusters (**Figure 1** and **Supplementary**
436 **Figure 2, 3 and 5**), normalised PC score could be used to attribute either *Sp* (low-diversity) or *Sk*
437 (high-diversity) ancestry to each window (summary statistics for each window are given in
438 **Supplementary Table 2)**. This was performed both for the subset of 57 samples (**Figure 1c**) and for
439 all 161 samples (**Supplementary Figure 6**). Using different window sizes (150, 100 and 50 SNPs
440 with overlap of 75, 50 and 25 respectively) yielded qualitatively the same results. Intermediate values
441 in PC1 (between 0.25 and 0.75) were only observed in few, sequential windows where samples

transitioned between clusters (**Supplementary Figure 3**). The only exception was sample JB1207, which we found to be diploid (for details see below).

Population structure after hybridization

To characterise the genome-wide distribution of ancestry components along the genome, we ran a hierarchical cluster analysis on the matrix containing ancestry information (*Sp* or *Sk*) for each window (columns) and strain (rows) using the R package *Pvclust 2.0.0* (Suzuki and Shimodaira, 2006). *Pvclust* includes a multiscale bootstrap resampling approach to calculate approximately unbiased probability values (p-values) for each cluster. We specified 1000 bootstraps using the Ward method and a Euclidian-based dissimilarity matrix. The analysis was run both for the whole genome (**Figure 1c**) and by chromosome (**Figure 1c**, **Supplementary Figure 9**).

Phylogenetic analysis of the mitochondrial genome

From the VCF file, we extracted mitochondrial variants for all 161 samples (Jeffares et al., 2015) and generated an alignment in *.fasta* format by substituting SNPs into the reference *S. pombe* 972 h− reference genome (ASM294v264) using the package *vcf2fasta* (https://github.com/JoseBlanca/vcf2fasta/, version Nov. 2015). We excluded variants in mitochondrial regions with SVs inferred from long reads. A maximum likelihood tree was calculated using *RaxML* (version 8.2.10-gcc-mpi) (Stamatakis, 2014) with default parameters, GTRGAMMAI approximation, final optimization with GTR + GAMMA + I and 1000 bootstraps. The final tree was visualised using *FigTree* 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) (**Supplementary Figure 10**).

Time of hybridization

Previous work (Jeffares et al., 2015) has shown that the time to the most recent common ancestor for 161 samples dates back to around 2300 years ago. This defines the maximum boundary for the time of hybridization. We used the theoretical model by Janzen et al., (2018) to infer the age of the initial hybridization event. The model predicts the number of ancestry blocks and junctions present in a hybrid individual as a function of time and effective population size (N_e). First, we obtained an estimate of N_e using the multiple sequential Markovian coalescent (MSMC). We constructed artificial diploid genomes from strains with consistent clustering by ancestry (**Figure 1c)** and estimated change in N_e as function across time using *MSMC* 2-2.0.0 (Schiffels and Durbin, 2014). In total we took four samples per group and produced diploid genomes in all possible six pairs for each group, except for one cluster that had only two samples (JB1205 and JB1206). Bootstraps were produced for each analysis, subsampling 25 genomic fragments per chromosome of 200 kb each. Resulting effective population size and time was scaled using reported mutation rate of $2 \cdot 10^{-10}$ mutations site^{-1} generation^{-1} (Farlow et al., 2015). Although it is difficult to be certain of the number of independent hybridization events, it is interesting to see that some clusters show similar demographic histories (**Supplementary**

Figure 24). Regardless of the demographic history in each cluster, long-term N_e as estimated by the harmonic mean ranged between $1{\cdot}10^5$ and $1{\cdot}10^9$. N_e of the near-pure ancestral *Sp* and *Sk* cluster was $7{\cdot}10^5$ and $9{\cdot}10^6$, respectively. These estimates of N_e are consistent with previous reports of $1{\cdot}10^7$ (Farlow et al., 2015).

We then used a customised R script with the ancestral component matrix to estimate the number of ancestry blocks (*Sp* or *Sk* clade) (**Supplementary Figure 12**). We used the R script from Janzen et al., (2018), and ran the model in each sample and chromosome using: $N_e = 1{\cdot}10^6$, r = number of genomic windows per chromosome, $h0 = 0.298$ (mean heterogenicity ($h0$) was estimated from the ancestral haplotype matrix) and $c = 7.1$, 5.6, and 4.1 respectively for chromosome I, II and III (values taken from Munz et al. (1989)) (**Supplementary Figure 13**). Given the large N_e, no changes in mean heterogenicity is expected over time after hybridization due to drift (the proportion of ancestral haplotypes *Sp* and *Sk* in hybrids, estimated as $2pq$, where p and q are the proportion of each ancestral clade in hybrids). Accordingly, results did not change within the range of the large N_e values. For this analysis, samples with proportion of admixture lower than 0.1 were excluded. The analysis was repeated with different windows sizes (200, 100 and 50 SNPs per window).

Demographic validation

To differentiate between possible historical scenarios underlying the observed pattern of ancestral divergence and introgression we performed individual based forward simulations using *SLiM 3.2.1* (Haller and Messer, 2019). These simulations were not intended to trace the "real" demographic history including estimates of split time, changes in effective population size, population structure after hybridization, or migration rates between subpopulations. Much rather, we contrasted two extremes of a continuum of possible scenarios: i) a scenario of ancestral divergence in isolation followed by a short, recent period of gene flow and ii) a scenario of divergence with continuous gene flow throughout the course of evolution. We asked the questions which of the simulated scenarios may better recover the empirically observed deep divergence of ancestry components and their distribution in blocks along the genome. We parametrized the simulations with estimates from the literature including $1{\cdot}10^6$ for N_e, a mutation rate of $2{\cdot}10^{-10}$ site^{-1} generation^{-1} (Farlow et al., 2015), an average recombination rate of $1{\cdot}10^6$ site^{-1} generation^{-1} (Munz et al., 1989), and a cloning rate of 0.95 generation^{-1} (equivalent to around 1 sexual cycle every 20 asexual generations). In order to reduce the computational effort, all parameters were scaled relative to a N_e of 1000 as suggested in the SLiM manual. Simulations were divided in two parts: 1) simulations of single 15 kb genomic windows corresponding to ~200 SNPs per window in the empirical data; 2) whole chromosome simulations with a fragment of 3 Mb corresponding to the size of *S. pombe* chromosome III. In both cases, we simulated divergence of two populations that were connected by different levels of symmetrical gene flow ranging from migration rates of 0 (complete isolation) to $1{\cdot}10^{-4}$ (equivalent to 100 migrants per

generation in a population of $1 \cdot 10^6$). To mimic the empirical difference in genetic diversity between populations, effective population sizes of the populations were set to $7 \cdot 10^5$ and $3 \cdot 10^5$, respectively. For the first part (single windows), a total of 100 replicate simulations were run for $30 \cdot N_e$ generations. Every 2 N_e generations 50 individuals were randomly sampled in total from both populations. At each time point, samples were processed in the same way as the empirical data per window estimating the percentage of variance explained by PC1 and D_{xy} between subgroups as inferred by normalized PC1 scores (**Supplementary Figure 7**). For the second part (whole chromosome), a total of 100 replicate simulations were run for $15 \cdot N_e$ generations after which a total of 50 genomes were randomly sampled from both populations. Simulated data were divided into genomic windows of 200 SNPs and processed in the same way as the empirical data (**Supplementary Figure 8**) for which the mean proportion of variance explained by PC1 was 64% and mean divergence between subpopulations (D_{xy}) was 0.0058. These values were only obtained in the simulations of a single genomic window under migration rates below $2 \cdot 10^{-5}$ (equivalent to 20 migrants per generation in a population of $1 \cdot 10^6$ individuals). However, this migration rate was not enough to produce recombinant blocks in whole chromosome simulations. Blocks were only observed when migration rate exceeded $1 \cdot 10^{-4}$ (100 migrants per generation). Yet, with this level of gene flow the proportion of variance explained by PC1 (around 25% in all simulations) and D_{xy} (below 0.0025 in all simulations) were significantly lower.

Phenotypic variation and reproductive isolation

We sourced phenotypic data of 229 phenotypic measurements in the 161 strains including amino acid quantification on liquid chromatography (aaconc), growth and stress on solid media (smgrowth), cell growth parameters and kinetics in liquid media (lmgrowth) and cell morphology (shape1 and shape2) from Jeffares et al. (2015). Data on reproductive isolation measured as the percentage of viable spores in pairs of crosses were compiled from Jeffares et al. (2015) and Marsellach (2017). A summary of all phenotypic measurements and reproductive data is provided in **Supplementary Table 4** and **5**, respectively.

First, we normalized each phenotypic trait y using rank-based transformation with the relationship normal.y = qnorm(rank (y) / (1 + length(y))). We then conducted PCA on normalized values of all phentoypic traits using the R package *missMDA* 1.12 (Josse and Husson, 2016). We estimated the number of dimensions for the principal component analysis by cross-validation, testing up to 30 PC components and imputing missing values. In addition to PCA decomposing variance across all traits, we examined the effect of admixture on each trait separately. Samples were divided into three discrete categories of admixture: two groups including samples with low admixture proportions (proportion of *Sp* or *Sk* clades higher than 0.9), and one for hybrid samples (proportion of *Sp* or *Sk* clades between 0.1 to 0.9). Significant differences in phenotypic distributions between groups were tested using *Tukey*

Honest Significant Differences as implemented in *Stats 3.4.2* (Team, 2014). **Supplementary Figure 14** shows the distribution of phenotypic values by admixture category for each trait. The number of traits with significant differences among groups was contrasted to values obtained by randomising admixture categories without replacement (permutations of the *Sp*, *Sk,* or *hybrid* category). Observed values were contrasted with distribution of the expected number of significant traits after running 10000 independent permutations (**Supplementary Figure 15)**.

Heritability

Heritability was estimated for all normalized traits using *LDAK* 5.94 (Speed et al., 2012), calculating independent kinship matrices derived from: 1) all SNPs and 2) ancestral haplotypes. Both SNPs and haplotype data were binary encoded (0 or 1). Jeffares et al. (2015) showed that heritability estimates between normalised and raw values are highly correlated ($r = 0.69$, p-value $\leq 2.2 \cdot 10^{-16}$). Heritability estimated with SNP values were strongly correlated with those from ancestral haplotypes ($r = 0.82$, p-value $\leq 2.2 \cdot 10^{-16}$). Heritability estimates and standard deviation for each trait for both SNPs and ancestral haplotypes are detailed in **Supplementary Table 6**.

Identification of ploidy changes

S. pombe is generally considered haploid under natural conditions. Yet, for two samples ancestry components did not separate on the principle component axis 1 (see above) for much of the genome. Instead, these samples were intermediate in PC1 score. A possible explanation is diploidisation of the two ancestral genomes. To establish the potential ploidy of samples, we called variants for all 161 samples using the Illumina data from Jeffares *et al* (2015) . Cleaned reads were mapped with *BWA* (version 0.7.17-r1188) in default settings and variants were called using *samtools* and *bcftools* (version 1.8). After filtering reads with a QUAL score > 25, the number of heterozygous sites per base per 20kb window were calculated. Additionally the nuclear content (C) as measured by Jeffares et al. (2015) (Supplementary Table S4 in Jeffares *et al* (2015)) were used to verify increased ploidy. Two samples showed high heterozygosity along the genome (JB1169 and JB1207) of which JB1207 for which data were available also showed a high C-value, suggesting that these samples are diploid (**Supplementary Figures 4 & 25)**. In JB1207, heterozygosity varies along the genome, with regions of high and low diversity. Sample JB1110 showed genomic content similar to JB1207, but did not show heterozygosity levels above that of haploid strains, suggesting the increase in genome content occurred by autoploidization.

High-weight genomic DNA extraction and whole genome sequencing

To obtain high weight gDNA for long-read sequencing, we grew strains from single colonies and cultured them in 200 mL liquid EMM at 32 °C shaking at 150 r.p.m. overnight. Standard media and growth conditions were used throughout this work (Hagan et al., 2016) with minor modifications: We

used standard liquid Edinburgh Minimal Medium (EMM; Per liter: Potassium Hydrogen Phthalate 3.0 g, Na $HPO_4 \cdot 2H_2O$ 2.76 g, NH_4Cl 5.0 g, D-glucose 20 g, $MgCl_2 \cdot 6H2O$ 1.05 g, $CaCl_2 \cdot 2H_2O$ 14.7 mg, KCl 1 g, Na_2SO_4 40 mg, Vitamin Stock ×1000 1.0 ml, Mineral Stock ×10,000 0.1 ml, supplemented with 100 mg l^{-1} adenine and 225 mg l^{-1} leucine) for the asexual growth. DNA extraction was performed with Genomic Tip 500/G or 100/G kits (Qiagen) following the manufacturer's instruction, but using Lallzyme MMX for lysis (Flor-Parra et al 2014, doi:10.1002/yea.2994). For each sample, 20 kb libraries were produced that were sequenced on one SMRT cell per library using the Pacific Biosciences RSII Technology Platform (PacBio®, CA). For a subset of eight samples, additional sequencing was performed using Oxford Nanopore (MinION). Sequencing was performed at SciLifeLab, Uppsala, Gene centre LMU, Munich and the Genomics & Bioinformatics Laboratory, University of York. We obtained on average 80x (SMRT) and 140x (nanopore) coverage for the nuclear genome for each sample (summary in **Supplementary Table 7**).

Additionally, 2.5 µg of the same DNA was delivered to the SNP&SEQ Technology Platform at the Uppsala Biomedical Centre (BMC), for Illumina sequencing. Libraries were prepared using the TruSeq PCRfree DNA library preparation kit (Illumina Inc.). Sequencing was performed on all samples pooled into a single lane, with cluster generation and 150 cycles paired-end sequencing on the HiSeqX system with v2.5 sequencing chemistry (Illumina Inc.). These data were used for draft genome polishing (see below).

De novo assembly of single-molecule read data

De novo genomes were assembled with *Canu* 1.5 (Koren et al., 2017) using default parameters. BridgeMapper from the *SMRT* 2.3.0 package was used to polish and subsequently assess the quality of genome assembly. Draft genomes were additionally polishing using short Illumina reads, running four rounds of read mapping to the draft genome with *BWA* 0.7.15 and polishing with *Pilon* 1.22 (Walker et al., 2014). Summary statistics of the final assembled genomes are found in **Supplementary Table 7**. *De novo* genomes were aligned to the reference genome using *MUMmer* 3.23 (Kurtz et al., 2004). Contigs were classified by reference chromosome to which they showed the highest degree of complementary. We used customised python scripts to identify and trim mitochondrial genomes.

Structural variant detection

Structural variants (SVs) were identified by a combination of a *de novo* and mapping approach. *De novo* genomes were aligned to the reference genome using *MUMmer*, and SVs were called using the function show-diff and the package *SVMU* 0.2beta (Khost et al., 2017). Then, raw long reads were mapped to the reference genome with *NGMLR* and genotypes were called using the package *Sniffles* (Sedlazeck et al., 2018). We implemented a new function within Sniffles "forced genotypes", which calls SVs by validating the mapping calls from an existing list of breaking points or SVs. This reports

the read support per variant even down to a single read. We forced genotypes using the list of *de novo* breaking points to generate a multi-sample VCF file. SVs were merged using the package *SURVIVOR* (Jeffares et al., 2017) option merge with a threshold of 1kbp and requiring the same type. In total, it resulted in a list of 1498 SVs with 892 in common between the mapping and *de-novo* approaches **(Supplementary Figure 16)**.

Within the 892 common variants we compared the accuracy of genotyping between sample by comparing genotypes obtained from *de novo* genomes and by mapping reads to reference genome. Additionally, we compared genotypes in samples sequenced with both PacBio and MinIon. In total we sequenced 8 samples with both technologies. We found high consistency for variants called with both sequencing technologies and observed that allele frequencies were highly correlated ($r = 0.98$, p-value $\leq 2.2 \times 10^{-16}$) (**Supplementary Figures 16 - 19**).
Only common SVs between the mapping and *de-novo* approach were considered, and variants with consistency below 50% were removed. We manually checked large SVs (larger than 10kb) by comparing the list of SVs with the alignment of the *de novo* genomes to the reference genome from MUMmer. This resulted in a final data set with 832 SVs (**Supplementary Table 8**).

Distribution of SVs around developmentally programmed DNA double-strand breaks (DSB)
We tested the association between DSB and SVs by comparing the physical genomic coordinates of the final list of SVs with DSB locations accessed from Fowler et al., (2014). Maintaining the same number of SVs per chromosome, we used a customized R script to randomise SV coordinates and measure the distances to the closest DSB. We counted the number of SVs present within different intervals of physical genetic distance ([0,500), [500, 1000), [1000, 2000), [2000, 4000), [4000, 10000), [10000, 20000), [20000, 30000) bp). Empirically observed values were contrasted with randomized distribution after running 10000 independent permutations. P-values of differences between randomization and observed values were obtained from the fraction of expected values higher than the observed value from the original data **(Supplementary Figures 20)**.

PCR validation of large SVs
To test the frequency of large inversions and rearrangements observed from long read data, we performed PCR verification over the breakpoints in the 57 non-clone samples. PCR was performed for both sides of the breakpoints, with a combination of one primer ‘outside’ of the inversion and both primers ‘inside’ the inversion (**Supplementary Figure 26**). PCR were performed on DNA using standard *Taq* polymerase, with annealing temperature at 59°C. The primers used, the coordinates in the reference and the expected amplicon length are given in **Supplementary Table 9**.

Distribution of structural variants in ancestral population – Two dimensional folded site frequency spectrum

We used the location of break points of SVs to identify whether a variant was located in the *Sp* or *Sk* genetic background in each sample. Ancestral haplotypes are difficult to infer in telomeric and centromeric regions given the low confidence in SNP calling in those regions, resulting in low percentage of variance explained by PC1. Thus SVs with break points in those regions were excluded from this analysis (19 SVs). SVs were grouped by ancestral group and allele frequencies were calculated for each ancestral population. We used these frequencies to build a two dimensional folded site frequency spectrum (2dSFS). In order to compare this 2dSFS, we repeated the analysis using SNP data from all 57 samples. Considering that the majority of identified SVs with long reads were transposable elements, we also made use of LRT insertion-deletion polymorphism (indels) inferred from short reads. For this additional data we produced a similar folded 2d SFS. LTR indel data were taken from Jeffares et al., (2015) and are listed in **Supplementary Table 11**.

Decay in linkage disequilibrium (LD)

To contrast LD between alleles from alternative ancestral groups, we calculated LD between all described genomic windows within chromosomes (**Supplementary Figure 22**). For this analysis only hybrid samples were considered (strains with admixture proportion higher than 0.1). For each pair of windows, we polarized windows by ancestry (at a threshold of > 0.7) and calculated standardized LD as the squared Pearson's correlation coefficient (R^2) (Hill and Robertson, 1968; Weir, 1979). This measurement takes into consideration difference in allele frequencies. The expected value of R^2 ($E(R^2)$) can be approximated by (Hill and Weir, 1988):

$$E(R^2) = \left(\frac{10 + C}{(2 + C) * (11 + C)}\right) * \left(1 + \frac{(3 + C) * (12 + 12C + C^2)}{n * (2 + C) * (11 + C)}\right)$$

Where C corresponds to product between the genetic distance (bp) and the population recombination rate (ρ) in n number of haplotype sampled. The population recombination rate was calculated as: $\rho = 4 * N_e * c$, where c is the recombination fraction between sites and Ne is the effective population size. We fitted a nonlinear model to obtain least squares estimates of ρ using a customized R script. The decay of LD with physical distance can be described with this model (Remington et al., 2001). LD values were grouped in three categories: i) comparison between windows with high proportion (*Sp*>0.7) of *Sp* ancestral group (*Sp*-*Sp*); ii) high proportion (*Sk*>0.7) of *Sk* ancestral group (*Sk*-*Sk*); and iii) high proportion of opposite ancestral groups (*Sp*-*Sk*). i) and ii) represent cases of positive ancestry disequilibrium, iii) will be denoted as negative ancestry disequilibrium. In order to reduce bias given the potential for secondary population structure (clusters in **Figure 1c**), the analysis was repeated using one random sample per cluster, resulting in similar general conclusions (**Supplementary Figure 27**).

Data availability

Nanopore, single-molecule real time sequencing data and de-novo genomes are available at NCBI Sequence Read Archive, BioProject ID PRJNA527756.

Acknowledgments

We thank Fidel Botero-Castro, Ana Catalán, Sebastian Höhna, Ulrich Knief, Claire Peart, Joshua Peñalba, Ricardo Pereira, Matthias Weissensteiner, (LMU Munich) and S. Lorena Ament-Velásquez (Uppsala University) for providing valuable intellectual input on the various analyses, sharing scripts and critically comment on the manuscript. We are further indebted to Bernadette Weissensteiner for extensive help with laboratory work and Saurabh Pophaly for bioinformatics support (LMU Munich). We further acknowledge support for data generation from the National Genomics Infrastructure, Uppsala, Sweden, the Gene Centre, Munich, Germany, Sally James and Peter Ashton from the Bioscience Technology Facility, Department of Biology, University of York, U.K, and James Chong, Department of Biology, University of York, U.K. The computational infrastructure was provided by the UPPMAX Next-Generation Sequencing Cluster and Storage (UPPNEX) project funded by the Knut and Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing and the York Advanced Research Computing Cluster (YARCC), University of York, U.K. This study was funded by LMU Munich to JW and NHGRI UM1 HG008898 to FS.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A.W., Parisod, C., Pfennig, K., Rice, A.M., Ritchie, M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf, J.B.W., Zinner, D., 2013. Hybridization and speciation. J. Evol. Biol. 26, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R.J., Barton, N.H., Good, J.M., 2016. Genomics of hybridization and its evolutionary consequences. Mol. Ecol. 25, 2325–2332. https://doi.org/10.1111/mec.13685

Brown, W.R.A., Liti, G., Rosa, C., James, S., Roberts, I., Robert, V., Jolly, N., Tang, W., Baumann, P., Green, C., Schlegel, K., Young, J., Hirchaud, F., Leek, S., Thomas, G., Blomberg, A., Warringer, J., 2011. A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. G3 GenesGenomesGenetics 1, 615–626. https://doi.org/10.1534/g3.111.001123

Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L., Ayroles, J.F., 2013. Genetic incompatibilities are widespread within species. Nature 504, 135–137. https://doi.org/10.1038/nature12678

Currall, B.B., Chiangmai, C., Talkowski, M.E., Morton, C.C., 2013. Mechanisms for Structural Variation in the Human Genome. Curr. Genet. Med. Rep. 1, 81–90. https://doi.org/10.1007/s40142-013-0012-8

Dowling, T. E., Smith, G. R. & Brown, W. M. Reproductive isolation and introgression between *Notropis cornutus* and *Notropis chrysocephalus* (family Cyprinidae): comparison of morphology, allozymes, and mitochondrial DNA. *Evolution.* **43,** 620–634 (1989).

Eroukhmanoff, F., Hermansen, J.S., Bailey, R.I., Sæther, S.A., Sætre, G.-P., 2013. Local adaptation within a hybrid species. Heredity 111, 286–292. https://doi.org/10.1038/hdy.2013.47

Faria, R., Navarro, A., 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol. Evol. 25, 660–669. https://doi.org/10.1016/j.tree.2010.07.008

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T.G., Nordborg, M., Lynch, M., 2015. The Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe. Genetics 201, 737–744. https://doi.org/10.1534/genetics.115.177329

Fawcett, J.A., Iida, T., Takuno, S., Sugino, R.P., Kado, T., Kugou, K., Mura, S., Kobayashi, T., Ohta, K., Nakayama, J., Innan, H., 2014. Population Genomics of the Fission Yeast Schizosaccharomyces pombe. PLOS ONE 9, e104241. https://doi.org/10.1371/journal.pone.0104241

Fisher, R.A., 1999. The genetical theory of natural selection: a complete variorum edition. Oxford University Press.

Fowler, K.R., Sasaki, M., Milman, N., Keeney, S., Smith, G.R., 2014. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. Genome Res. 24, 1650–1664. https://doi.org/10.1101/gr.172122.114

Hagan, I.M., Carr, A.M., Grallert, A., Nurse, P., 2016. Fission yeast: a laboratory manual. Cold Spring Harbor Laboratory Press.

Haller, B.C., Messer, P.W., 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Mol. Biol. Evol. 36, 632–637. https://doi.org/10.1093/molbev/msy228

Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231. https://doi.org/10.1007/BF01245622

Hill, W.G., Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33, 54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hoffman, C.S., Wood, V., Fantes, P.A., 2015. An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. Genetics 201, 403–423. https://doi.org/10.1534/genetics.115.181503

Hoffmann, A.A., Rieseberg, L.H., 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? Annu. Rev. Ecol. Evol. Syst. 39, 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Hu, W., Jiang, Z.-D., Suo, F., Zheng, J.-X., He, W.-Z., Du, L.-L., 2017. A large gene family in fission yeast encodes spore killers that subvert Mendel's law. eLife 6. https://doi.org/10.7554/eLife.26057

Janzen, T., Nolte, A. W. & Traulsen, A. The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites: breakdown of ancestry blocks after hybridization. *Evolution.* **72,** 735–750 (2018).

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J., Dasmahapatra, K.K., Joron, M., 2018. Supergene Evolution Triggered by the

Introgression of a Chromosomal Inversion. Curr. Biol. 28, 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072
Jeffares, D.C., 2018. The natural diversity and ecology of fission yeast. Yeast 35, 253–260. https://doi.org/10.1002/yea.3293
Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., Sedlazeck, F.J., 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat. Commun. 8, 14061. https://doi.org/10.1038/ncomms14061
Jeffares, D.C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., Marsellach, F.X., Iqbal, Z., Lau, W., Cheng, T.M.K., Pracana, R., Mülleder, M., Lawson, J.L.D., Chessel, A., Bala, S., Hellenthal, G., O'Fallon, B., Keane, T., Simpson, J.T., Bischof, L., Tomiczek, B., Bitton, D.A., Sideri, T., Codlin, S., Hellberg, J.E.E.U., van Trigt, L., Jeffery, L., Li, J.-J., Atkinson, S., Thodberg, M., Febrer, M., McLay, K., Drou, N., Brown, W., Hayles, J., Salas, R.E.C., Ralser, M., Maniatis, N., Balding, D.J., Balloux, F., Durbin, R., Bähler, J., 2015. The genomic and phenotypic diversity of Schizosaccharomyces pombe. Nat. Genet. 47, 235–241. https://doi.org/10.1038/ng.3215
Josse, J., Husson, F., 2016. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. J. Stat. Softw. 70. https://doi.org/10.18637/jss.v070.i01
Kelly, F.D., Levin, H.L., 2005. The evolution of transposons in *Schizosaccharomyces pombe*. Cytogenet. Genome Res. 110, 566–574. https://doi.org/10.1159/000084990
Khost, D.E., Eickbush, D.G., Larracuente, A.M., 2017. Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. Genome Res. 27, 709–721. https://doi.org/10.1101/gr.213512.116
Kondrat'eva, V.I., Naumov, G.I., 2001. The Phenomenon of Spore Killing in Schizosaccharomyces pombe Hybrids. Dokl. Biol. Sci. 379, 385–388. https://doi.org/10.1023/A:1011624918673
Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116
Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., Karlionova, N., Pinchuk, P., Verkuil, Y.I., Kitaysky, A.S., Wingfield, J.C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D.B., Burke, T., 2016. A supergene determines highly divergent male reproductive morphs in the ruff. Nat. Genet. 48, 79–83.
Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12.
Lamichhaney, S. *et al.* Rapid hybrid speciation in Darwin's finches. *Science* eaao4593 (2017).
Leducq, J.-B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J.-P., Samani, P., Sylvester, K., Hittinger, C.T., Bell, G., Landry, C.R., 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. Nat. Microbiol. 1, 15003. https://doi.org/10.1038/nmicrobiol.2015.3
Mallet, J., 2007. Hybrid speciation. Nature 446, 279–283. https://doi.org/10.1038/nature05706
Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229–237.
Marsellach, X. A non-genetic meiotic repair program inferred from spore survival values in fission yeast wild isolates: a clue for an epigenetic ratchet-like model of ageing? (2017).

McDonald, M.J., Rice, D.P., Desai, M.M., 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature 531, 233–236. https://doi.org/10.1038/nature17143
Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8, ncomms14363. https://doi.org/10.1038/ncomms14363
Muller, H.J., 1932. Some genetic aspects of sex. Am. Nat. 66, 118–138.
Munz, P., Wolf, K., Kohli, J., Leupold, U., 1989. Genetics overview. Mol. Biol. Fission Yeast 1–30.
Naumov, G.I., Kondratieva, V.I., Naumova, E.S., 2015. Hybrid sterility of the yeast Schizosaccharomyces pombe: Genetic genus and many species in statu nascendi? Microbiology 84, 159–169. https://doi.org/10.1134/S0026261715010099
Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. 76, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269
Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, J.A., Soltis, P.S., Villa-Machío, I., 2017. Is homoploid hybrid speciation that rare? An empiricist's view. Heredity 118, 513–516. https://doi.org/10.1038/hdy.2017.7
Nieuwenhuis, B.P.S., Tusso, S., Bjerling, P., Stångberg, J., Wolf, J.B.W., Immler, S., 2018. Repeated evolution of self-compatibility for reproductive assurance. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-04054-6
Nolte, A.W., Sheets, H.D., 2005. Shape based assignment tests suggest transgressive phenotypes in natural sculpin hybrids (Teleostei, Scorpaeniformes, Cottidae). Front. Zool. 2, 11. https://doi.org/10.1186/1742-9994-2-11
Nuckolls, N.L., Bravo Núñez, M.A., Eickbush, M.T., Young, J.M., Lange, J.J., Yu, J.S., Smith, G.R., Jaspersen, S.L., Malik, H.S., Zanders, S.E., 2017. wtf genes are prolific dual poison-antidote meiotic drivers. eLife 6. https://doi.org/10.7554/eLife.26033
Ortiz-Barrientos, D., Engelstädter, J., Rieseberg, L.H., 2016. Recombination Rate Evolution and the Origin of Species. Trends Ecol. Evol. 31, 226–236. https://doi.org/10.1016/j.tree.2015.12.016
Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* n/a-n/a (2016).
Pennisi, E., 2016. A shortcut to a species. Science 354, 818–818. https://doi.org/10.1126/science.354.6314.818
Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G., Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5
Pfennig, K.S., Kelly, A.L., Pierce, A.A., 2016. Hybridization as a facilitator of species range expansion. Proc. R. Soc. B Biol. Sci. 283, 20161329. https://doi.org/10.1098/rspb.2016.1329
Poelstra, J.W., Vijay, N., Bossu, Christen, Lantz, Henrik, Ryll, Bettina, Müller, Inge, Baglione, Vittorio, Unneberg, Per, Wikelski, Martin, Grabherr, Manfred, Wolf, Jochen B. W., 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344, 1410–1414.
Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S., 2001. Structure of linkage

887 disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci.
888 98, 11479–11484. https://doi.org/10.1073/pnas.201394398
889 Runemark, A. *et al.* Variation and constraints in hybrid genome formation. *Nat. Ecol. Evol.*
890 (2018).
891 Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N.,
892 Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day
893 humans. Nature 507, 354–357. https://doi.org/10.1038/nature12961
894 Schiffels, S., Durbin, R., 2014. Inferring human population size and separation history from
895 multiple genome sequences. Nat. Genet. 46, 919–925. https://doi.org/10.1038/ng.3015
896 Schumer, M., Cui, R., Powell, D.L., Rosenthal, G.G., Andolfatto, P., 2016. Ancient
897 hybridization and genomic stabilization in a swordtail fish. Mol. Ecol. 25, 2661–2679.
898 https://doi.org/10.1111/mec.13602
899 Schumer, M., Rosenthal, G. G. & Andolfatto, P. How common is homoploid hybrid
900 speciation?: Perspective. *Evolution* **68,** 1553–1560 (2014).
901 Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A.,
902 Schatz, M.C., 2018. Accurate detection of complex structural variations using single-
903 molecule sequencing. Nat. Methods 15, 461–468. https://doi.org/10.1038/s41592-018-
904 0001-7
905 Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207.
906 https://doi.org/10.1016/j.tree.2004.01.003
907 Singh, G., Klar, A.J., 2002. The 2.1-kb inverted repeat DNA sequences flank the mat2, 3
908 silent region in two species of Schizosaccharomyces and are involved in epigenetic
909 silencing in Schizosaccharomyces pombe. Genetics 162, 591–602.
910 Soltis, P.S., Soltis, D.E., 2009. The Role of Hybridization in Plant Speciation. Annu. Rev.
911 Plant Biol. 60, 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039
912 Speed, D., Hemani, G., Johnson, M.R., Balding, D.J., 2012. Improved Heritability Estimation
913 from Genome-wide SNPs. Am. J. Hum. Genet. 91, 1011–1021.
914 https://doi.org/10.1016/j.ajhg.2012.10.010
915 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
916 large phylogenies. Bioinformatics 30, 1312–1313.
917 https://doi.org/10.1093/bioinformatics/btu033
918 Steenkamp, E.T., Wingfield, M.J., McTaggart, A.R., Wingfield, B.D., 2018. Fungal species
919 and their boundaries matter – Definitions, mechanisms and practical implications.
920 Fungal Biol. Rev. 32, 104–116. https://doi.org/10.1016/j.fbr.2017.11.002
921 Stukenbrock, E.H., 2016. The Role of Hybridization in the Evolution and Emergence of New
922 Fungal Plant Pathogens. Phytopathology 106, 104–112.
923 https://doi.org/10.1094/PHYTO-08-15-0184-RVW
924 Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in
925 hierarchical clustering. Bioinformatics 22, 1540–1542.
926 https://doi.org/10.1093/bioinformatics/btl117
927 Tajima, F., 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA
928 Polymorphism. Genetics 123, 585–595.
929 Taylor, D.J., Hebert, P.D., 1993. Habitat-dependent hybrid parentage and differential
930 introgression between neighboringly sympatric Daphnia species. Proc. Natl. Acad. Sci.
931 90, 7079–7083. https://doi.org/10.1073/pnas.90.15.7079
932 Team, R.C., 2014. R: A language and environment for statistical computing.
933 Teresa Avelar, A., Perfeito, L., Gordo, I., Godinho Ferreira, M., 2013. Genome architecture is
934 a selectable trait that can be maintained by antagonistic pleiotropy. Nat. Commun. 4,
935 2235. https://doi.org/10.1038/ncomms3235

936 The Heliconius Genome Consortium, Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D.,
937 Davey, J.W., Whibley, A., Nadeau, N.J., Zimin, A.V., Hughes, D.S.T., Ferguson, L.C.,
938 Martin, S.H., Salazar, C., Lewis, J.J., Adler, S., Ahn, S.-J., Baker, D.A., Baxter, S.W.,
939 Chamberlain, N.L., Chauhan, R., Counterman, B.A., Dalmay, T., Gilbert, L.E.,
940 Gordon, K., Heckel, D.G., Hines, H.M., Hoff, K.J., Holland, P.W.H., Jacquin-Joly, E.,
941 Jiggins, F.M., Jones, R.T., Kapan, D.D., Kersey, P., Lamas, G., Lawson, D., Mapleson,
942 D., Maroja, L.S., Martin, A., Moxon, S., Palmer, W.J., Papa, R., Papanicolaou, A.,
943 Pauchet, Y., Ray, D.A., Rosser, N., Salzberg, S.L., Supple, M.A., Surridge, A.,
944 Tenger-Trolander, A., Vogel, H., Wilkinson, P.A., Wilson, D., Yorke, J.A., Yuan, F.,
945 Balmuth, A.L., Eland, C., Gharbi, K., Thomson, M., Gibbs, R.A., Han, Y., Jayaseelan,
946 J.C., Kovar, C., Mathew, T., Muzny, D.M., Ongeri, F., Pu, L.-L., Qu, J., Thornton,
947 R.L., Worley, K.C., Wu, Y.-Q., Linares, M., Blaxter, M.L., ffrench-Constant, R.H.,
948 Joron, M., Kronforst, M.R., Mullen, S.P., Reed, R.D., Scherer, S.E., Richards, S.,
949 Mallet, J., Owen McMillan, W., Jiggins, C.D., 2012. Butterfly genome reveals
950 promiscuous exchange of mimicry adaptations among species. Nature 487, 94–98.
951 https://doi.org/10.1038/nature11041
952 Turner, L.M., Harr, B., 2014. Genome-wide mapping in a house mouse hybrid zone reveals
953 hybrid sterility loci and Dobzhansky-Muller interactions. eLife 3, e02504.
954 https://doi.org/10.7554/eLife.02504
955 Twyford, A.D., Streisfeld, M.A., Lowry, D.B., Friedman, J., 2015. Genomic studies on the
956 nature of species: adaptation and speciation in Mimulus. Mol. Ecol. 24, 2601–2609.
957 https://doi.org/10.1111/mec.13190
958 Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P., Wolf,
959 J.B.W., 2016. Evolution of heterogeneous genome differentiation across multiple
960 contact zones in a crow species complex. Nat. Commun. 7, 13195.
961 https://doi.org/10.1038/ncomms13195
962 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A.,
963 Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for
964 Comprehensive Microbial Variant Detection and Genome Assembly Improvement.
965 PLoS ONE 9, e112963. https://doi.org/10.1371/journal.pone.0112963
966 Watterson, G.A., 1975. On the number of segregating sites in genetical models without
967 recombination. Theor. Popul. Biol. 7, 256–276. https://doi.org/10.1016/0040-
968 5809(75)90020-9
969 Weir, B.S., 1979. Inferences about Linkage Disequilibrium. Biometrics 35, 235.
970 https://doi.org/10.2307/2529947
971 Wellenreuther, M., Bernatchez, L., 2018. Eco-Evolutionary Genomics of Chromosomal
972 Inversions. Trends Ecol. Evol. 33, 427–440. https://doi.org/10.1016/j.tree.2018.04.002
973 Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J.,
974 Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D.,
975 Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis,
976 P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J.,
977 Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K.,
978 James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P.,
979 Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S.,
980 Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D.,
981 Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K.,
982 Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward,
983 J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E.,
984 Rieger, M., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Düsterhöft, A., Fritzc,
985 C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H.,

Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dréano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sánchez, M., del Rey, F., Benito, J., Domínguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., 2003. Erratum: corrigendum: The genome sequence of Schizosaccharomyces pombe. Nature 421, 94–94. https://doi.org/10.1038/nature01203

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dréano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sánchez, M., del Rey, F., Benito, J., Domínguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerrutti, L., Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., 2002. Erratum: The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880. https://doi.org/10.1038/nature724

Zanders, S.E., Eickbush, M.T., Yu, J.S., Kang, J.-W., Fowler, K.R., Smith, G.R., Malik, H.S., 2014. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. eLife 3, e02630. https://doi.org/10.7554/eLife.02630

Zheng, X., Gogarten, S.M., Lawrence, M., Stilp, A., Conomos, M.P., Weir, B.S., Laurie, C., Levine, D., 2017. SeqArray—a storage-efficient high-performance data format for WGS variant calls. Bioinformatics 33, 2251–2257. https://doi.org/10.1093/bioinformatics/btx145

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Figure legends

Figure 1 | Distribution of *Sp* (red) and *Sk* (blue) ancestry blocks along the *S. pombe* genome. (a) Example of principal component analysis (PCA) of a representative genomic window in chromosome I (top) and the whole mitochondrial DNA (bottom). Samples fall into two major clades, *Sp* (red square) and *Sk* (blue square). The proportion of variance explained by PC1 and PC2 is indicated on the axis labels. Additional examples are found in Supplementary Figure 1 (b) Proportion of variance explained by PC1 (black line) and PC2 (grey line) for each genomic window along the genome. Centromeres are indicated with red bars. Note the drop in proximity to centromeres and telomeres where genotype quality is significantly reduced. (c) Heatmap for one representative of 57 near-clonal groups indicating ancestry along the genome (right panel). Samples are organised according to a hierarchical clustering, grouping samples based on ancestral block distribution (left dendrogram). Colours on the tips of the cladogram represent cluster membership by chromosome (see **Supplementary Figure 9)**. Samples changing clustering group between chromosomes are shown in grey. (d) Estimate of D_{xy} between ancestral groups and genetic diversity (π) within the *Sp* (red) and *Sk* clade (blue) along the genome.

Figure 2 | Population genetic summary statistics. (a) Proportion of *Sp* (red) and *Sk* (blue) ancestry across all 57 samples along the genome. (b) Tajima's D differentiated by *Sp* (red) and *Sk* (blue) ancestry and pooled across all samples irrespective of ancestry (grey line). Genomic regions previously identified under purifying selection (Fawcett et al., 2014) are shown with black triangles. Reported active meiotic drives (Zanders et al., 2014; Hu et al., 2017; Nuckolls et al., 2017) are indicated by yellow triangles. The third panel shows the difference between ancestry specific Tajima's D and the estimate from the pooled samples.

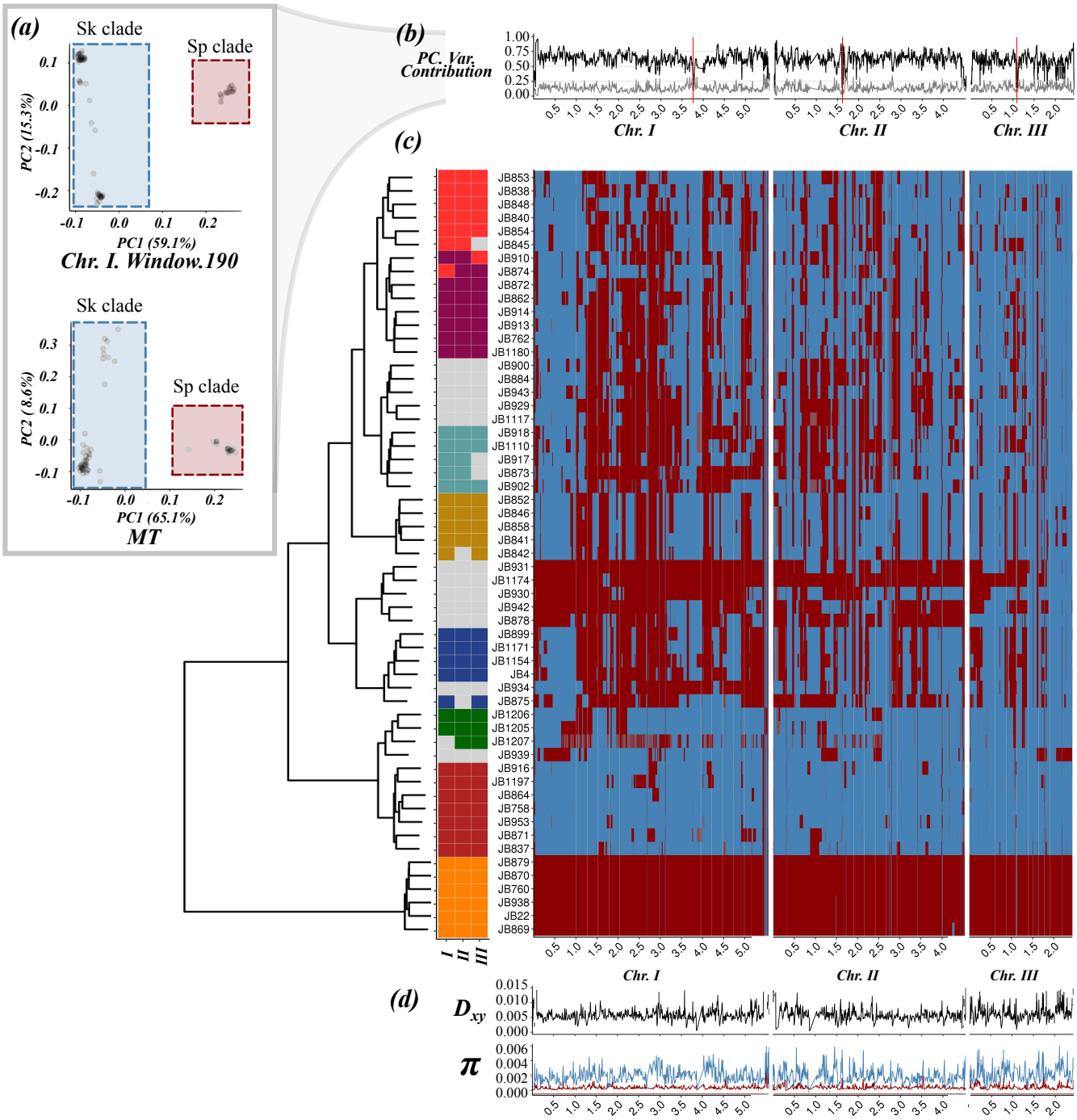
Figure 3 | Inferred evolutionary history of contemporary *S. pombe* strains. An ancestral population diverged into two major clades, *Sp* (red) and *Sk* (blue) since approximately 2300 years ago (Jeffares et al., 2015). Recurrent hybridization upon secondary contact initiated around 20-60 sexual outcrossing generations ago resulted in admixed genomes with a range of admixture proportions (bottom) prevailing today.

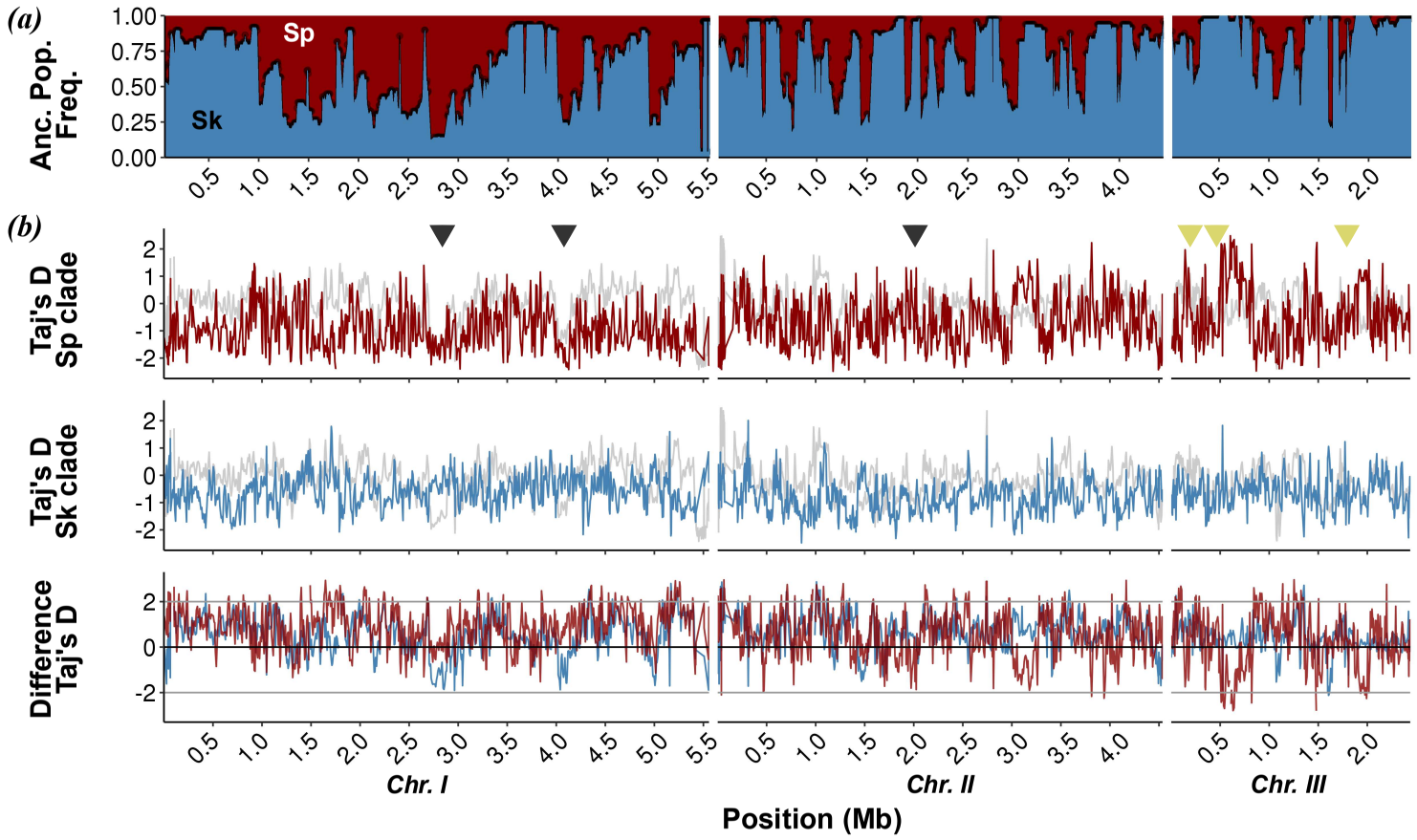
Figure 4 | Ancestry explains variation in phenotype and reproductive isolation. (a) PCA of normalized phenotypic variation across 228 traits. The proportion of variance explained by PC1 and PC2 is indicated on the axis labels. Admixed samples (dots) are coloured coded by ancestry proportion (cf. **Figure 3**) ranging from pure *Sp* (red triangle) to pure *Sk* (blue triangle) ancestry. (b) Phenotypic distribution of example traits separated by the degree of admixture: admixed samples are shown in grey, pure ancestral *Sp* and *Sk* samples are shown in red and blue respectively. The number of traits

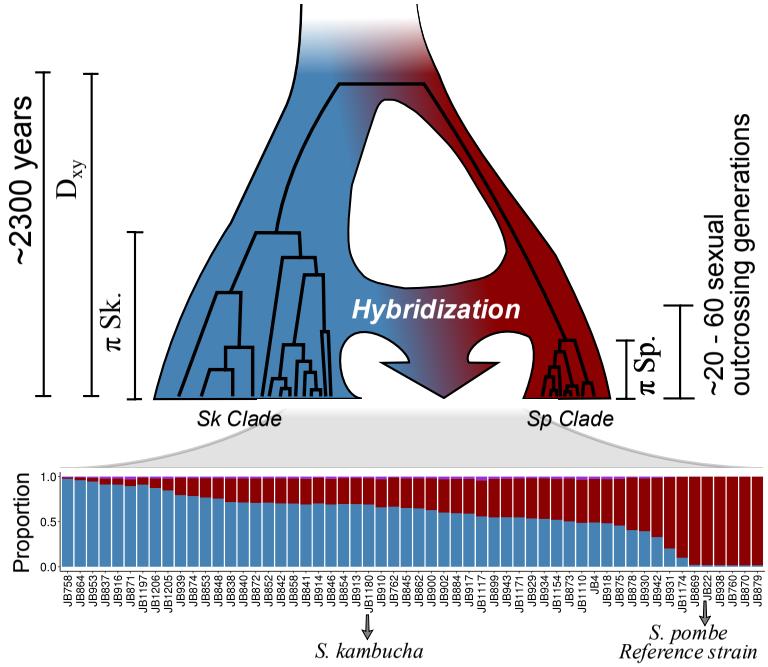
corresponding to a dominant, additive and transgressive genetic architecture is indicated on the right hand side (c) Comparison of heritability estimates of all 228 traits based on 172,935 SNPs (abscissa) and on 1925 genomic windows polarized by ancestry (ordinate). Colours indicate statistical significance. *NS*: heritability values not significantly different from zero, *AncHap:* significant only using ancestral blocks, *SNPs:* significant only using SNPs, *SB:* significant in both analyses. Diagonal (slope=1) added as reference. (d) Histogram of the difference between heritability estimates using SNPs and ancestry components for all 228 traits. (e) Correlation between the difference in ancestry proportions between two strains (cf. **Figure 2**) and spore viability of the cross. Red box shows samples with low spore viability but high genetic similarity.

Figure 5 | Characterization of structural variation based on long-read, real-time sequencing. (a) Schematic representation of the three chromosomes in different strains displaying SVs larger than 10kb relative to reference genome JB22 (left panel). Chromosome arms are differentiated by colour; orientation is indicated with arrows relative to the reference; black bars represent centromeres. In the second panel, additional SVs, their type and ID of the corresponding strain are illustrated in brackets. (b) Size distribution of SVs below 10 kb. Colours indicate the type of SV. (c) Distribution of SV density along the genome. Black bars represent centromeres. (d) Two-dimensional, folded site frequency spectrum between inferred ancestral populations for all SVs, SNPs and LTR INDELs. Numbers and colours show the percentage of the total number of variants in each category. Variants with low frequency in both populations are shown in the blue box. Variants highly differentiated between populations are show in red boxes with total in the upper right box. Fills with percentage lower than 0.01 are empty.
Abbreviations: DEL: deletion; DUP: duplication; INS: insertion; INV: inversion

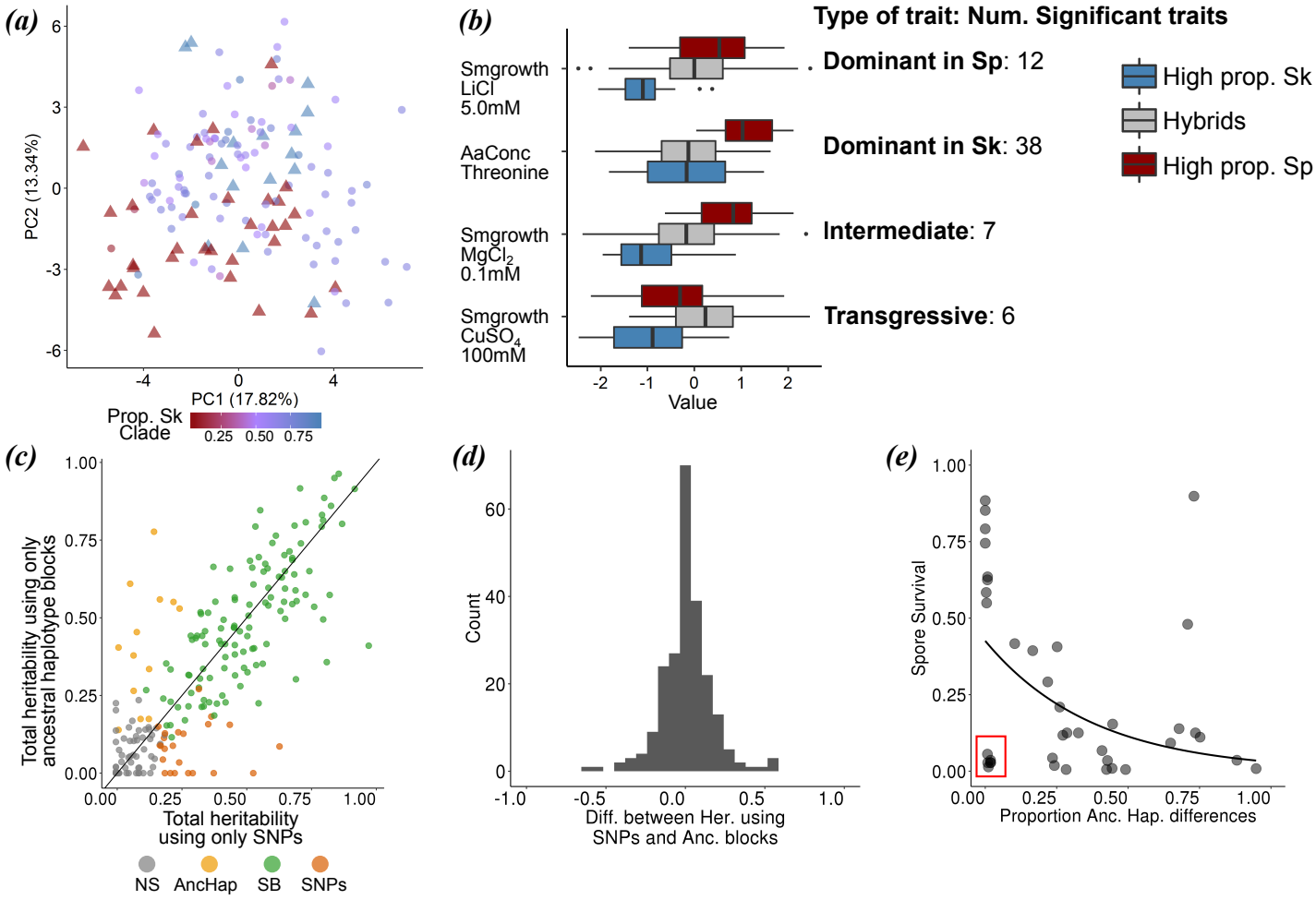
Figure 6 | Decay in linkage disequilibrium (LD) with genetic distance. Relationship between LD (R^2) and physical distance is depicted for each chromosome. Black points represent values for each window pair comparison. Lines show non-lineal regression model based on Hill & Weir (1988) and Remington et al. (2001). LD estimates were divided into three categories representing comparison between windows of shared ancestry (*Sp*-*Sp* or *Sk*-*Sk*) reflecting positive ancestry disequilibrium (AD) or of opposite ancestry (*Sk*-*Sp*) reflecting negative AD.

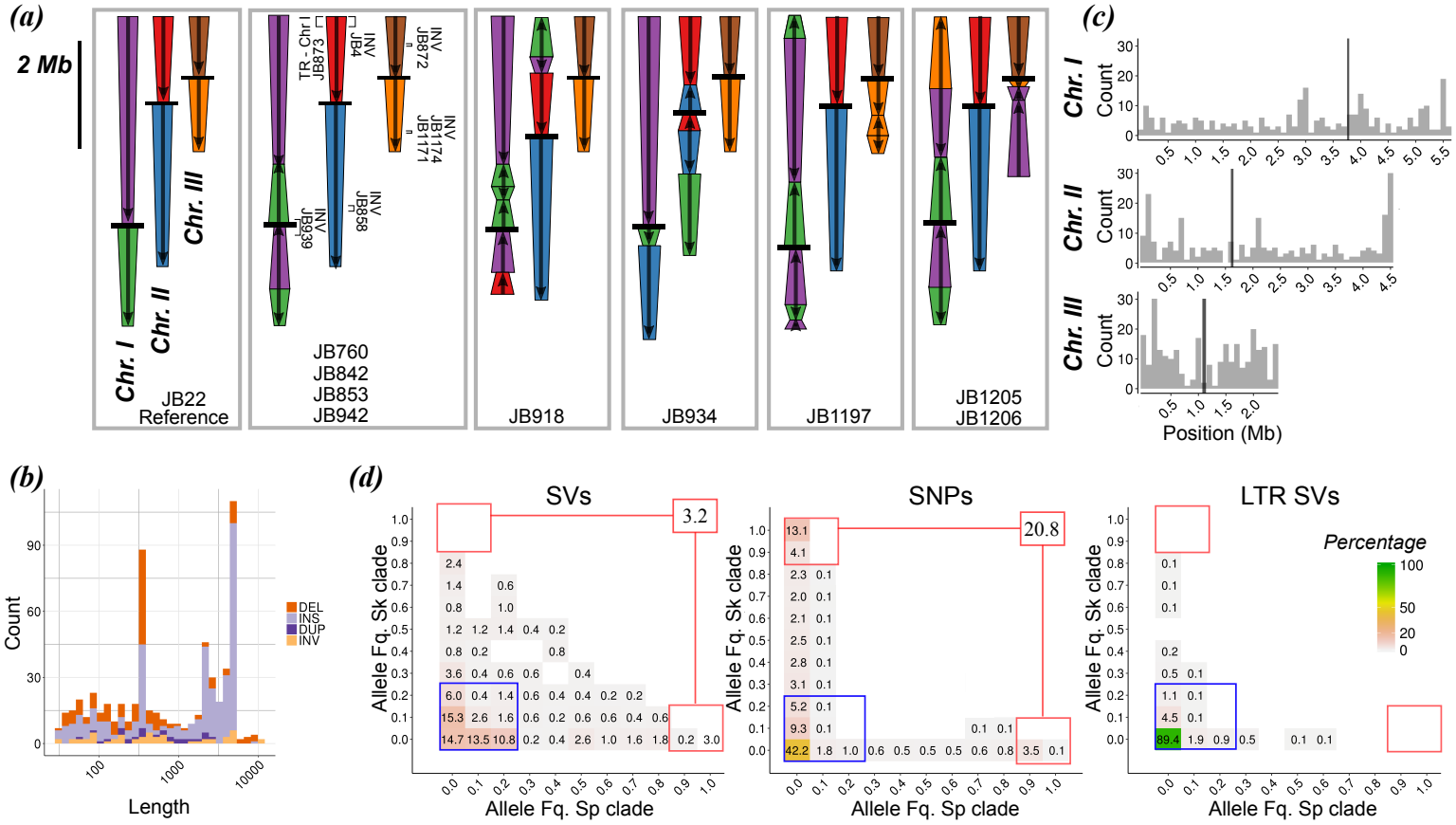

(a)
Sk clade
Sp clade
PC2 (15.3%)
PC1 (59.1%)
Chr. I. Window.190
Sk clade
Sp clade
PC2 (8.6%)
PC1 (65.1%)
MT
(b)
PC. Var. Contribution
Chr. I
Chr. II
Chr. III
(c)
JB853
JB838
JB848
JB840
JB854
JB845
JB910
JB874
JB872
JB862
JB914
JB913
JB762
JB1180
JB900
JB884
JB943
JB929
JB1117
JB918
JB1110
JB917
JB873
JB902
JB852
JB846
JB858
JB841
JB842
JB931
JB1174
JB930
JB942
JB878
JB899
JB1171
JB1154
JB4
JB934
JB875
JB1206
JB1205
JB1207
JB939
JB916
JB1197
JB864
JB758
JB953
JB871
JB837
JB879
JB870
JB760
JB938
JB22
JB869
I
II
III
Chr. I
Chr. II
Chr. III
(d)
Dxy
π

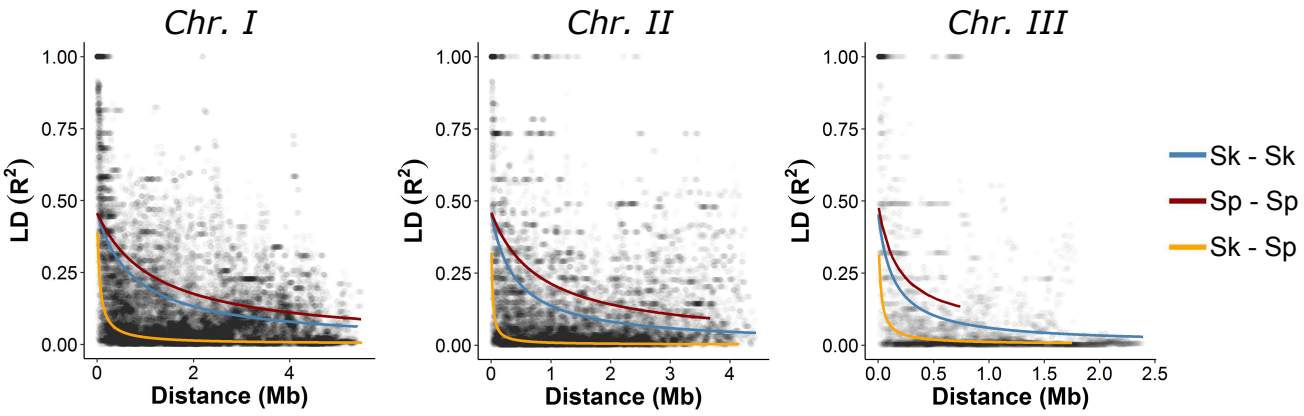

(a)
Anc. Pop. Freq.
1.00
0.75
0.50
0.25
0.00
Sp
Sk
(b)
Taj's D Sp clade
Taj's D Sk clade
Difference Taj's D
2
1
0
-1
-2
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
Chr. I
Chr. II
Chr. III
Position (Mb)


~2300 years
D_{xy}
π Sk.
Hybridization
π Sp.
~20 - 60 sexual outcrossing generations
Sk Clade
Sp Clade

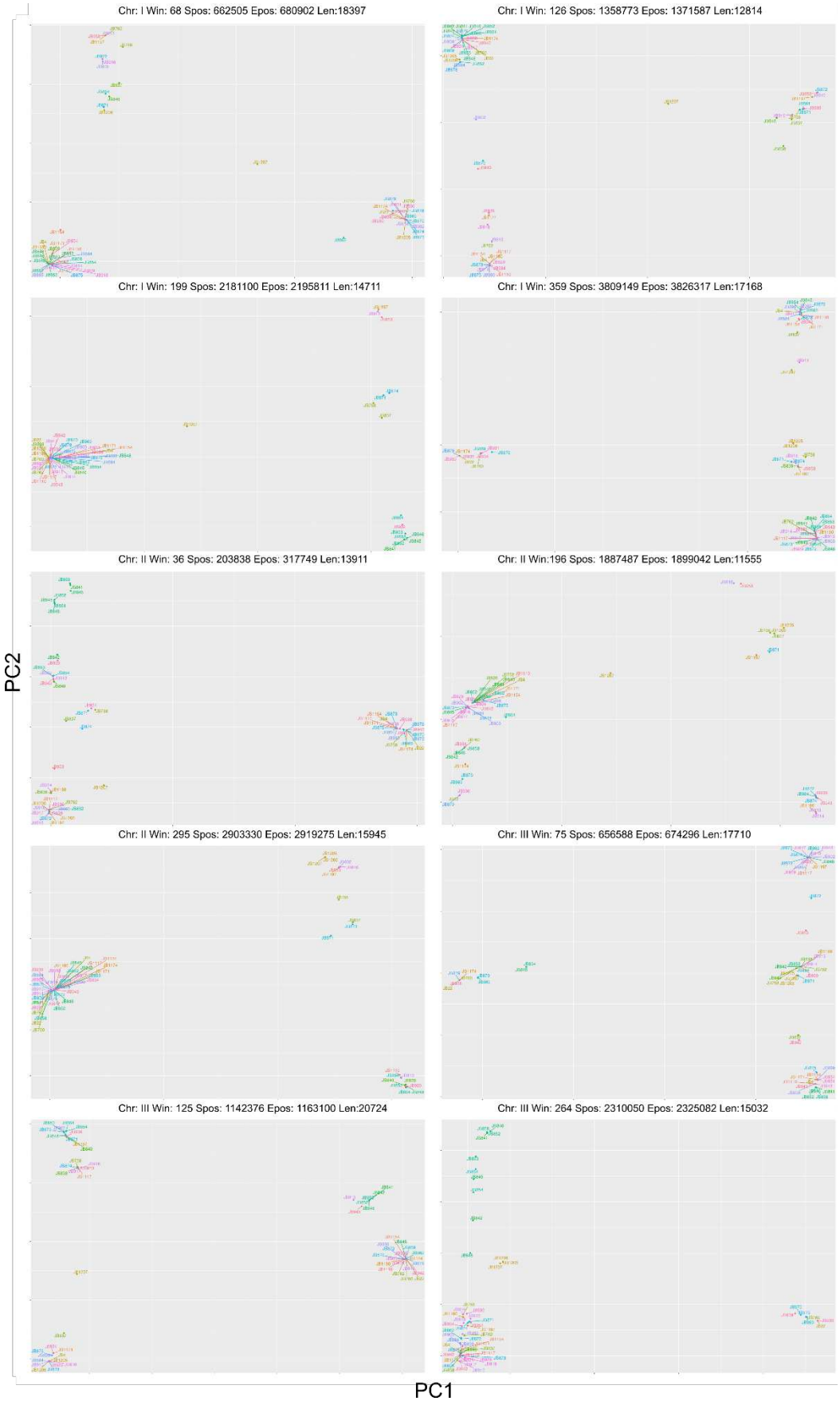
Proportion
1.0
0.5
0.0
JB758 JB864 JB953 JB837 JB916 JB871 JB1157 JB1206 JB1205 JB939 JB874 JB853 JB848 JB838 JB840 JB872 JB842 JB858 JB841 JB914 JB846 JB854 JB913 JB1160 JB910 JB752 JB845 JB862 JB900 JB902 JB884 JB917 JB1117 JB839 JB943 JB1171 JB929 JB934 JB1154 JB873 JB1110 JB4 JB918 JB875 JB878 JB930 JB942 JB931 JB1174 JB869 JB22 JB938 JB760 JB870 JB879
S. kambucha
S. pombe
Reference strain


(a)
PC2 (13.34%)
PC1 (17.82%)
Prop. Sk Clade
0.25 0.50 0.75
(b)
Type of trait: Num. Significant traits
Dominant in Sp: 12
Dominant in Sk: 38
Intermediate: 7
Transgressive: 6
High prop. Sk
Hybrids
High prop. Sp
Smgrowth LiCl 5.0mM
AaConc Threonine
Smgrowth MgCl2 0.1mM
Smgrowth CuSO4 100mM
Value
(c)
Total heritability using only ancestral haplotype blocks
Total heritability using only SNPs
NS AncHap SB SNPs
(d)
Count
Diff. between Her. using SNPs and Anc. blocks
(e)
Spore Survival
Proportion Anc. Hap. differences

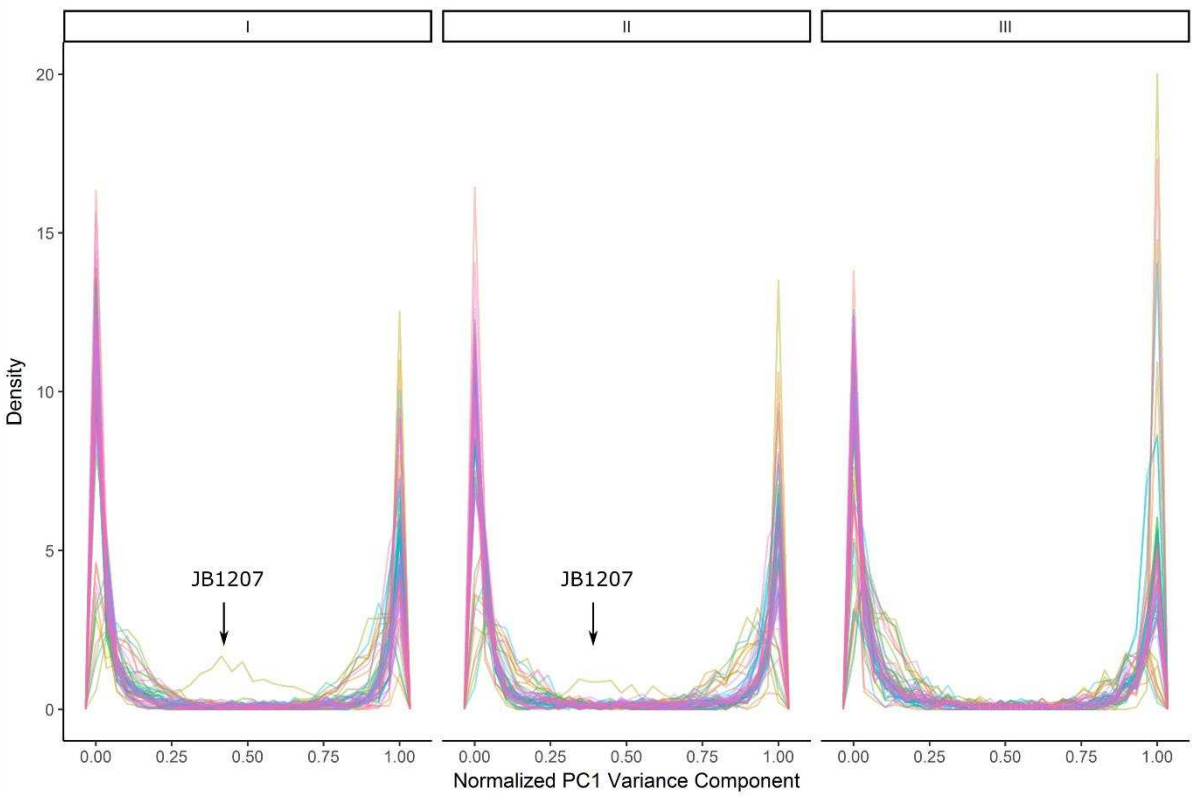

(a)
2 Mb
Chr. I
Chr. II
Chr. III
JB22
Reference
TR - Chr I
JB873
INV
JB84
INV
JB872
INV
JB1174
JB1171
INV
JB858
INV
JB939
JB760
JB842
JB853
JB942
JB918
JB934
JB1197
JB1205
JB1206
(c)
Chr. I
Chr. II
Chr. III
Count
Position (Mb)
(b)
Count
Length
DEL
INS
DUP
INV
(d)
SVs
SNPs
LTR SVs
Allele Fq. Sk clade
Allele Fq. Sp clade
Percentage
3.2
20.8


Chr. I
Chr. II
Chr. III
LD (R^2)
1.00
0.75
0.50
0.25
0.00
0
2
4
1
3
0.0
0.5
1.0
1.5
2.0
2.5
Distance (Mb)
Sk - Sk
Sp - Sp
Sk - Sp

Supplementary Figures


Chr: I Win: 68 Spos: 662505 Epos: 680902 Len:18397
Chr: I Win: 126 Spos: 1358773 Epos: 1371587 Len:12814
Chr: I Win: 199 Spos: 2181100 Epos: 2195811 Len:14711
Chr: I Win: 359 Spos: 3809149 Epos: 3826317 Len:17168
Chr: II Win: 36 Spos: 203838 Epos: 317749 Len:13911
Chr: II Win:196 Spos: 1887487 Epos: 1899042 Len:11555
Chr: II Win: 295 Spos: 2903330 Epos: 2919275 Len:15945
Chr: III Win: 75 Spos: 656588 Epos: 674296 Len:17710
Chr: III Win: 125 Spos: 1142376 Epos: 1163100 Len:20724
Chr: III Win: 264 Spos: 2310050 Epos: 2325082 Len:15032
PC2
PC1

Supplementary Figure 1: Example of principal component analysis (PCA). Each panel is an example PCA of different genomic windows. The chromosome (Chr.), window number (Win), start and end position (Spos and Epos) and length of the window in bp (Len) are shown in the header for each panel. Samples are differentiated by colours.



Supplementary Figure 2: Normalised PC1 distribution. Density plot of Normalised PC1 values shown separately for each of the three chromosomes. Each line represent one sample. The sample JB1207 is highlighted as it is the only sample with a high proportion of intermediate values.

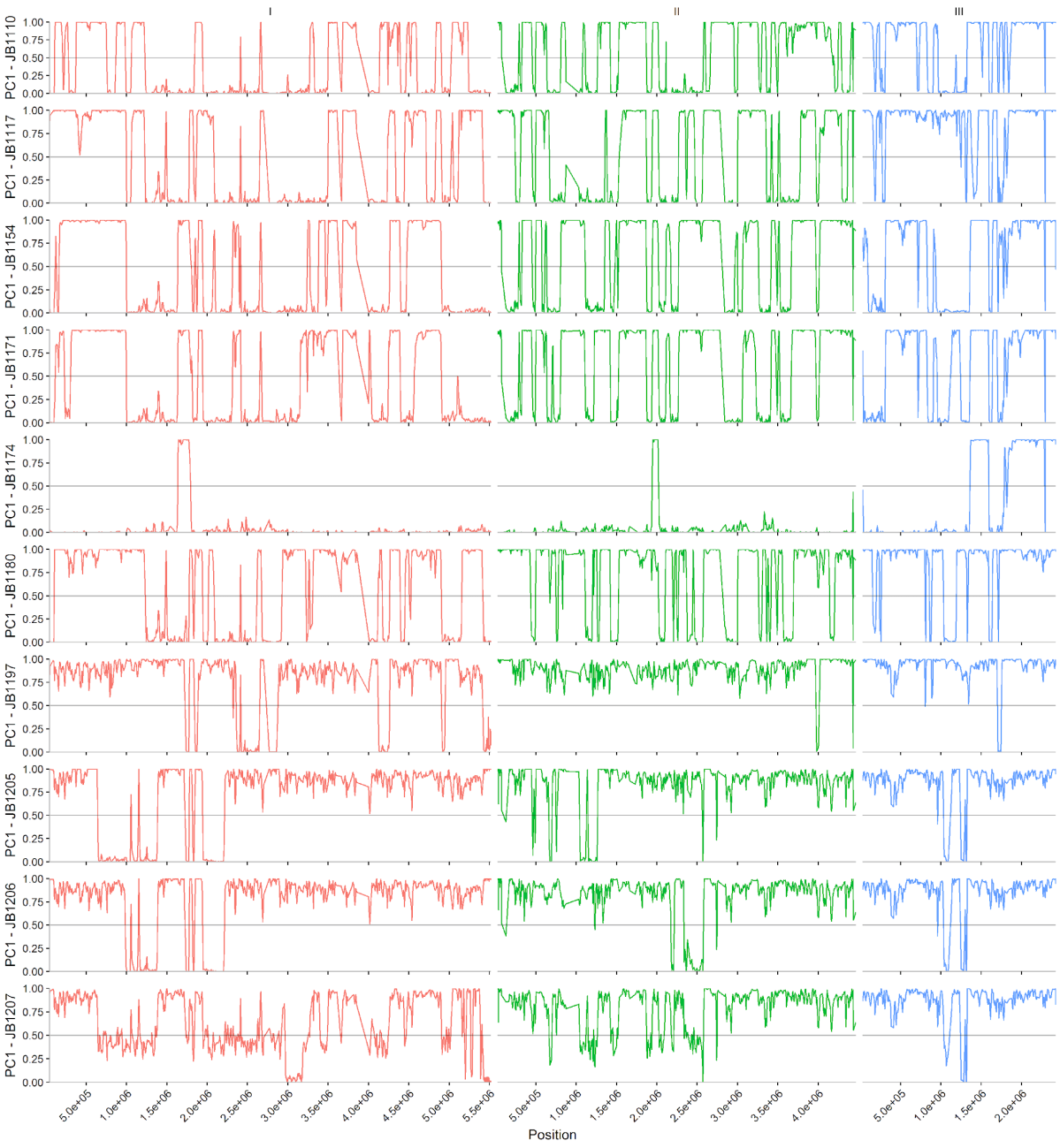


Figure continues on next page.

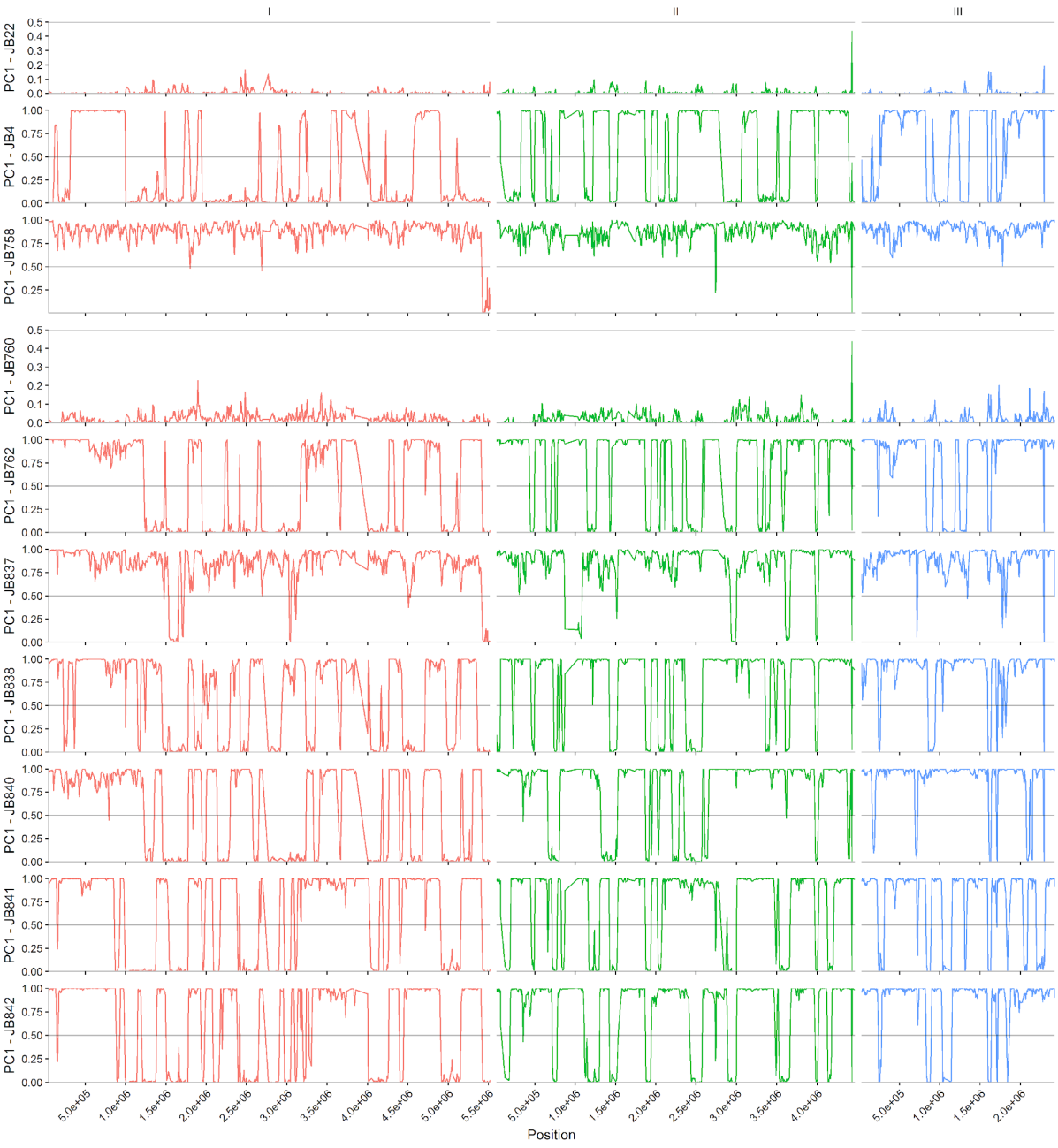


Figure continues on next page.

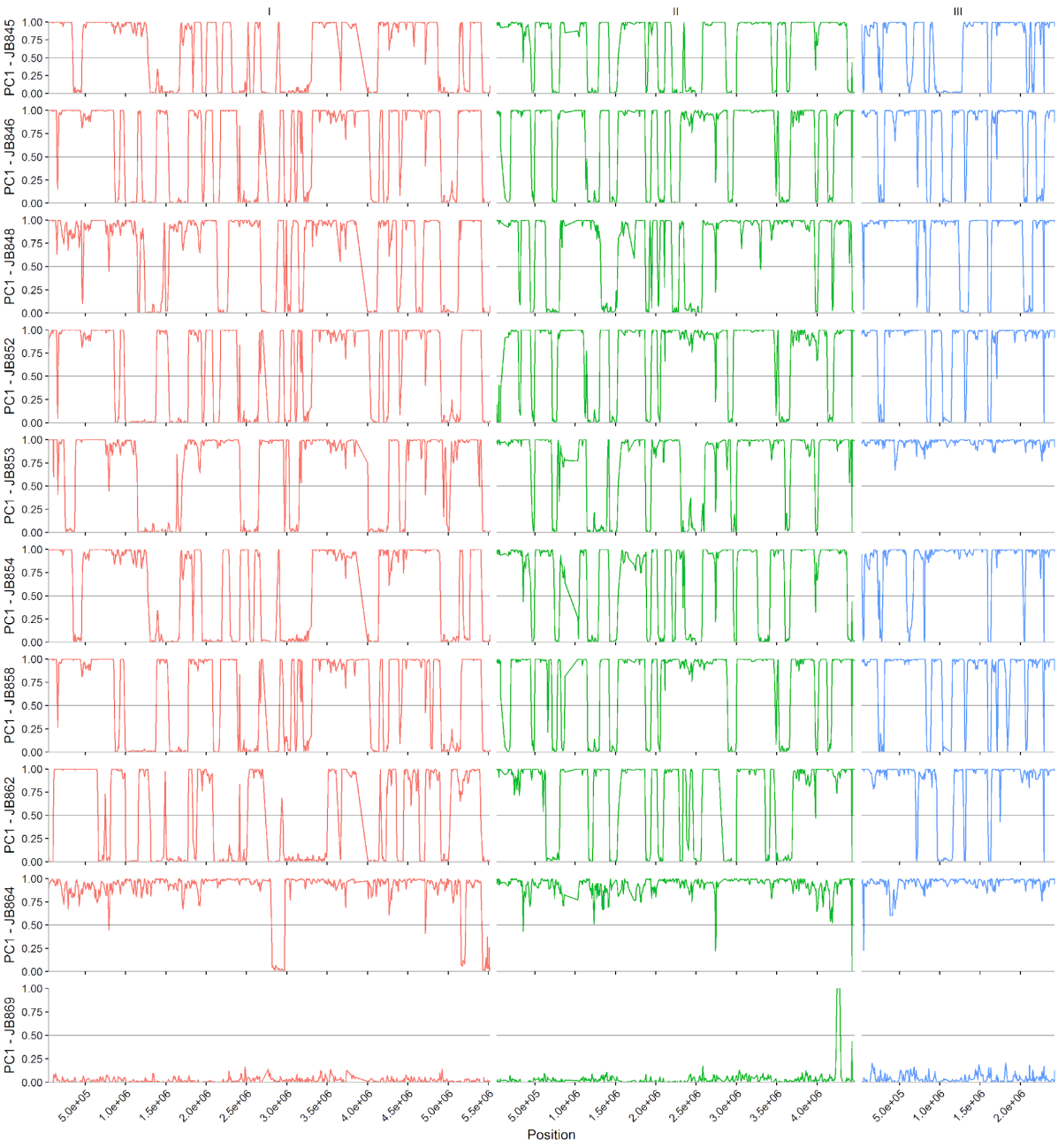


Figure continues on next page.

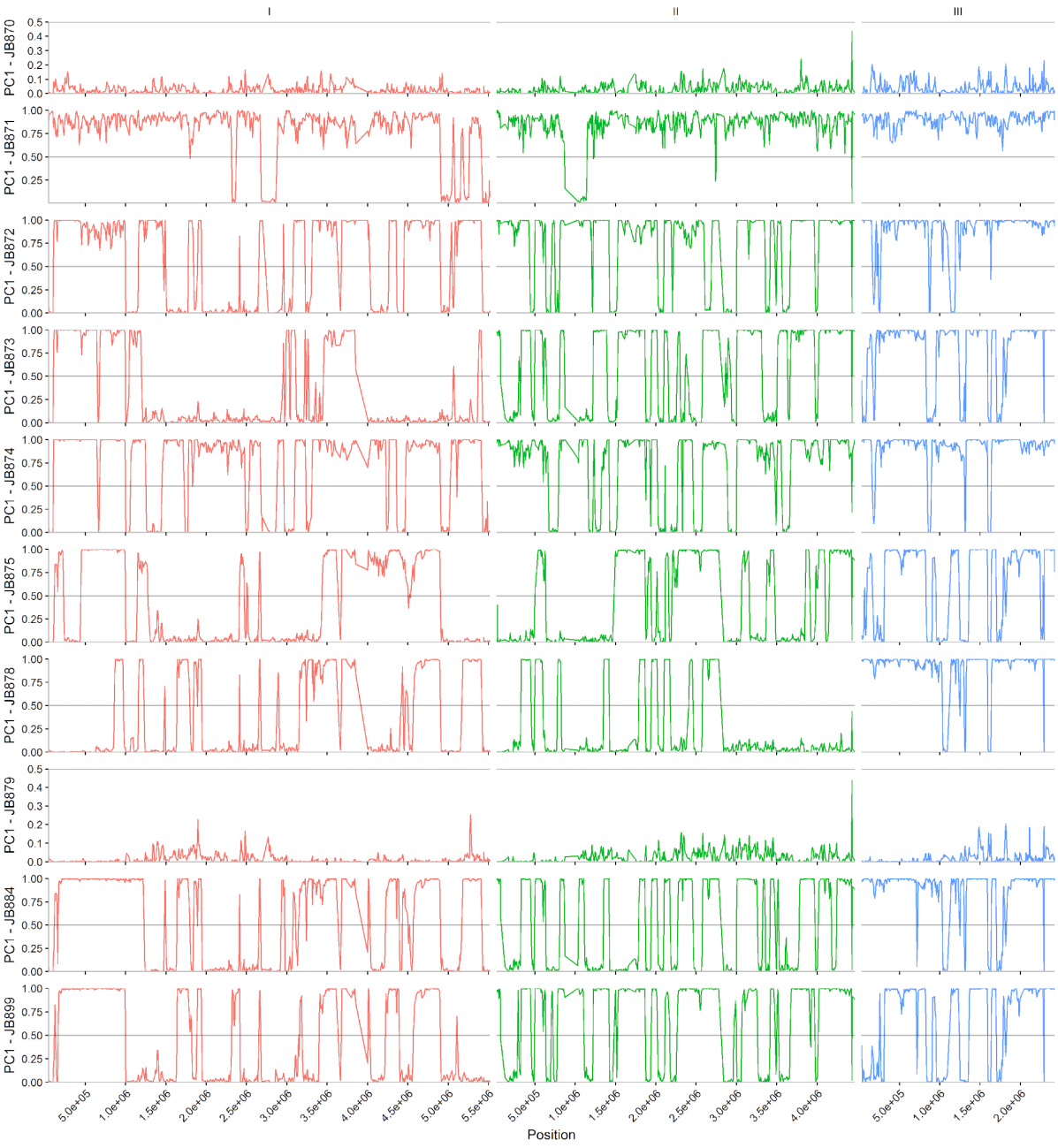


Figure continues on next page.

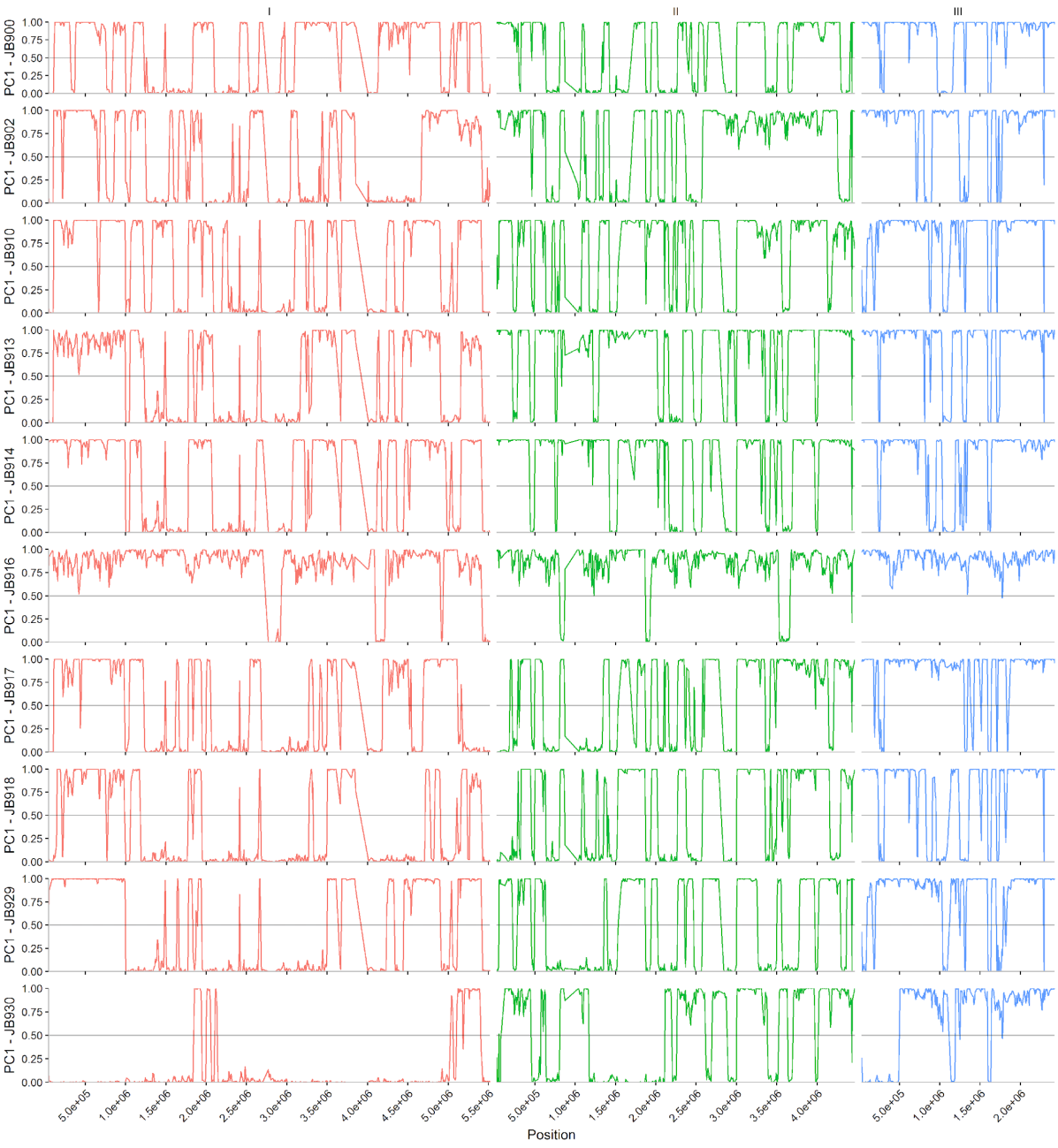
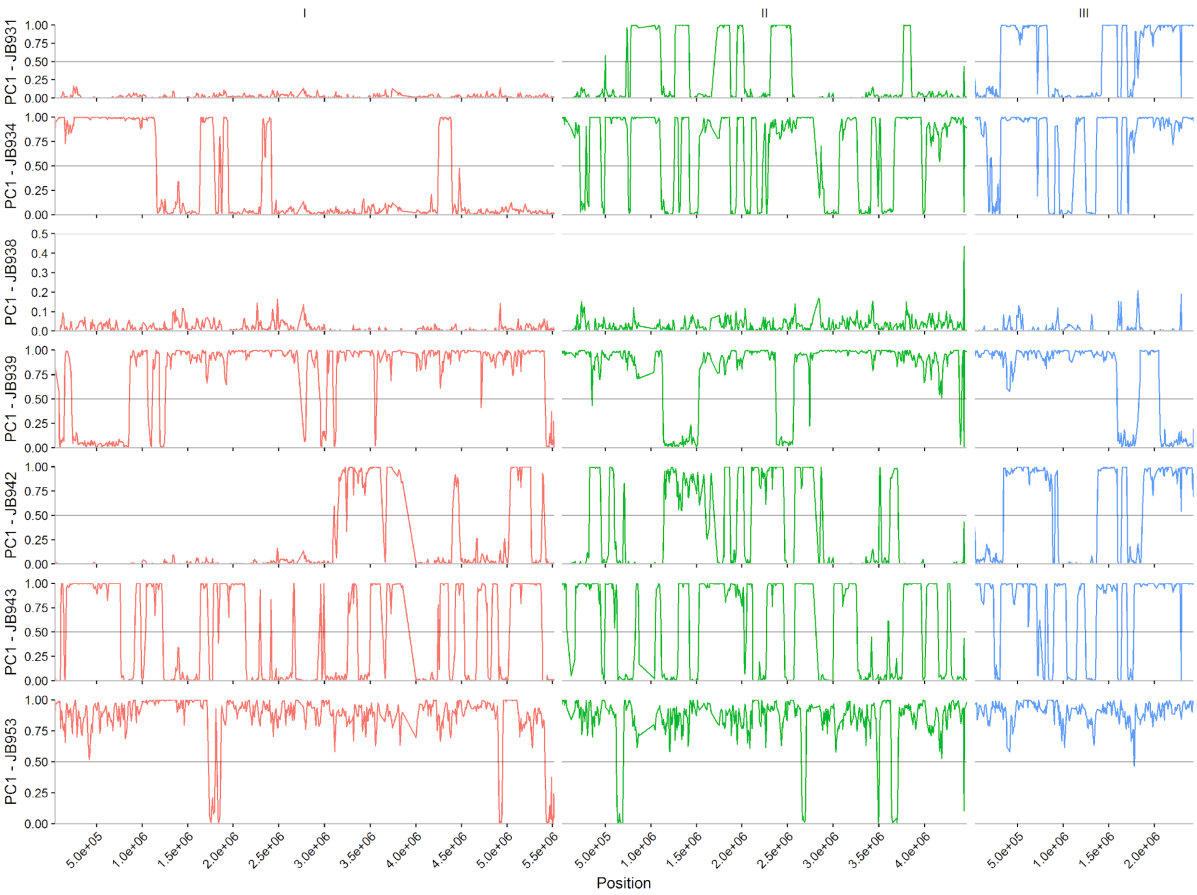
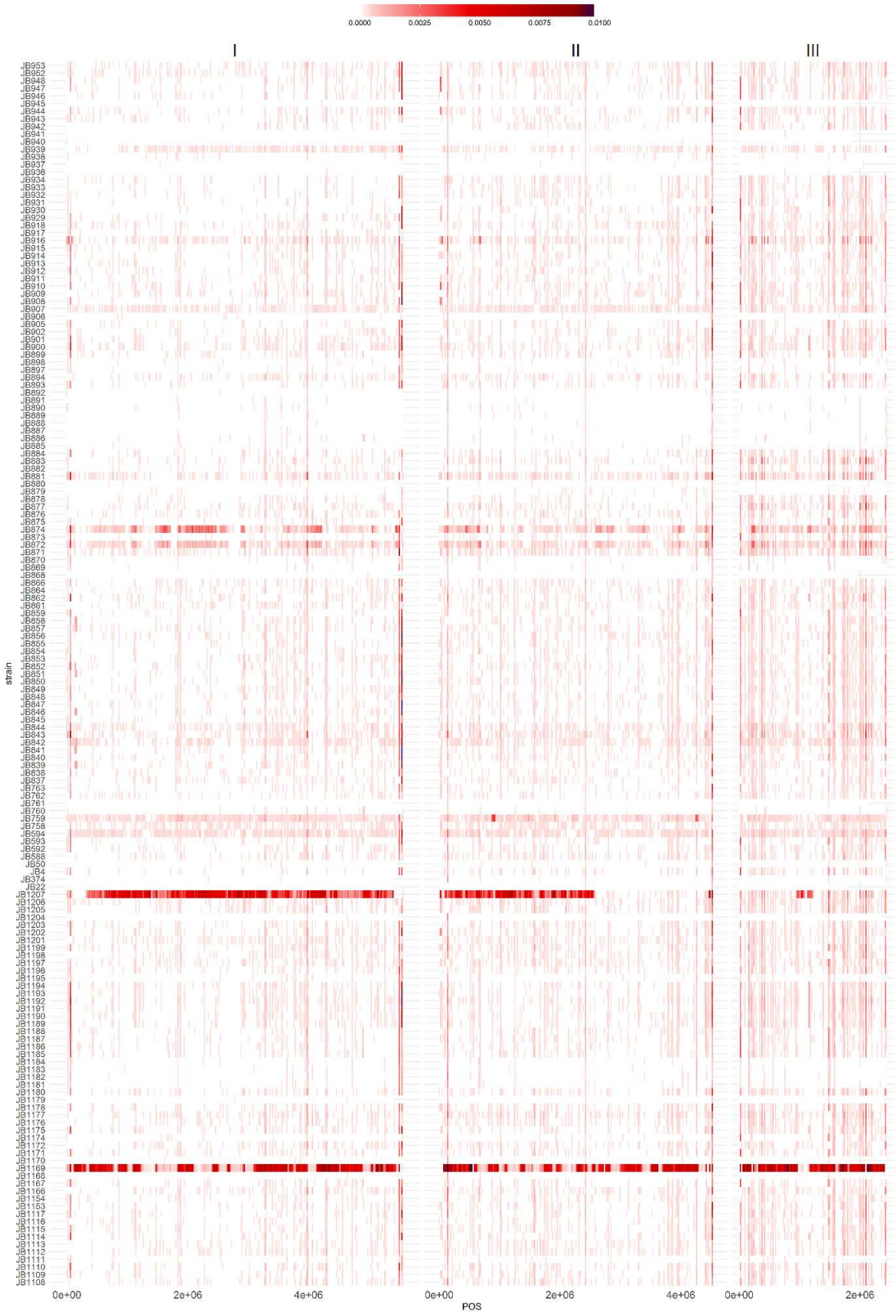


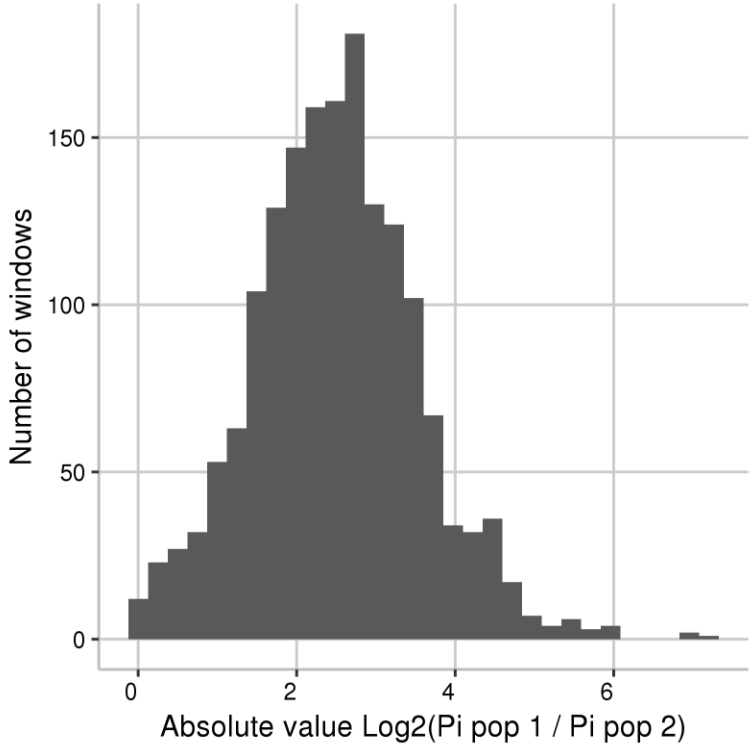
Figure continues on next page.



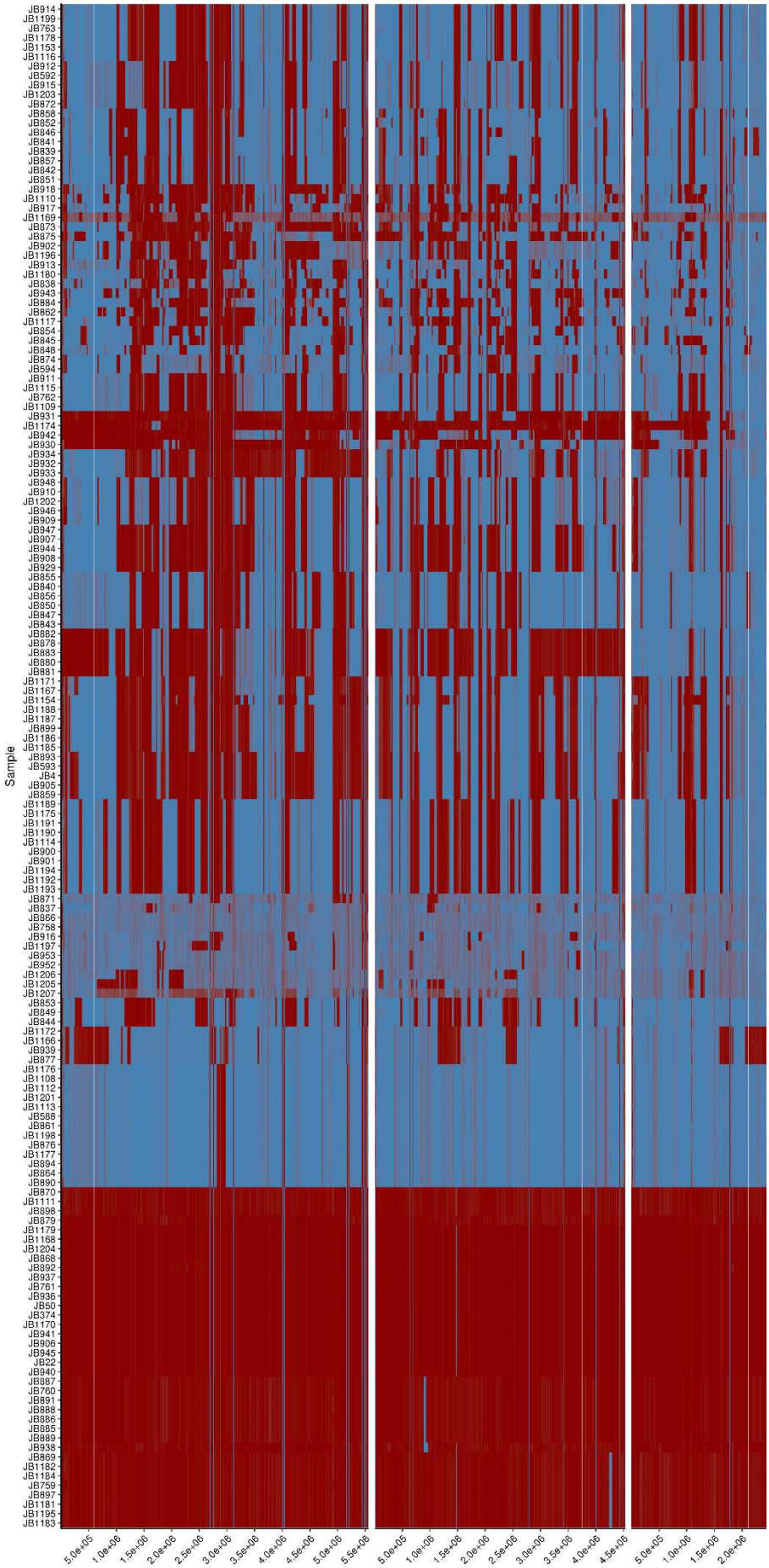
Supplementary Figure 3: Distribution of normalised PC1 values alone the genome. Chromosomes are indicated by colour (chr1 = red; chr2 = green, chr3 = blue). Normalised PC1 values were polarized based on genetic diversity as described in the methods section, values close to 0 or 1 represent genomic regions with the *Sp* or *Sk* haplotype, respectively.



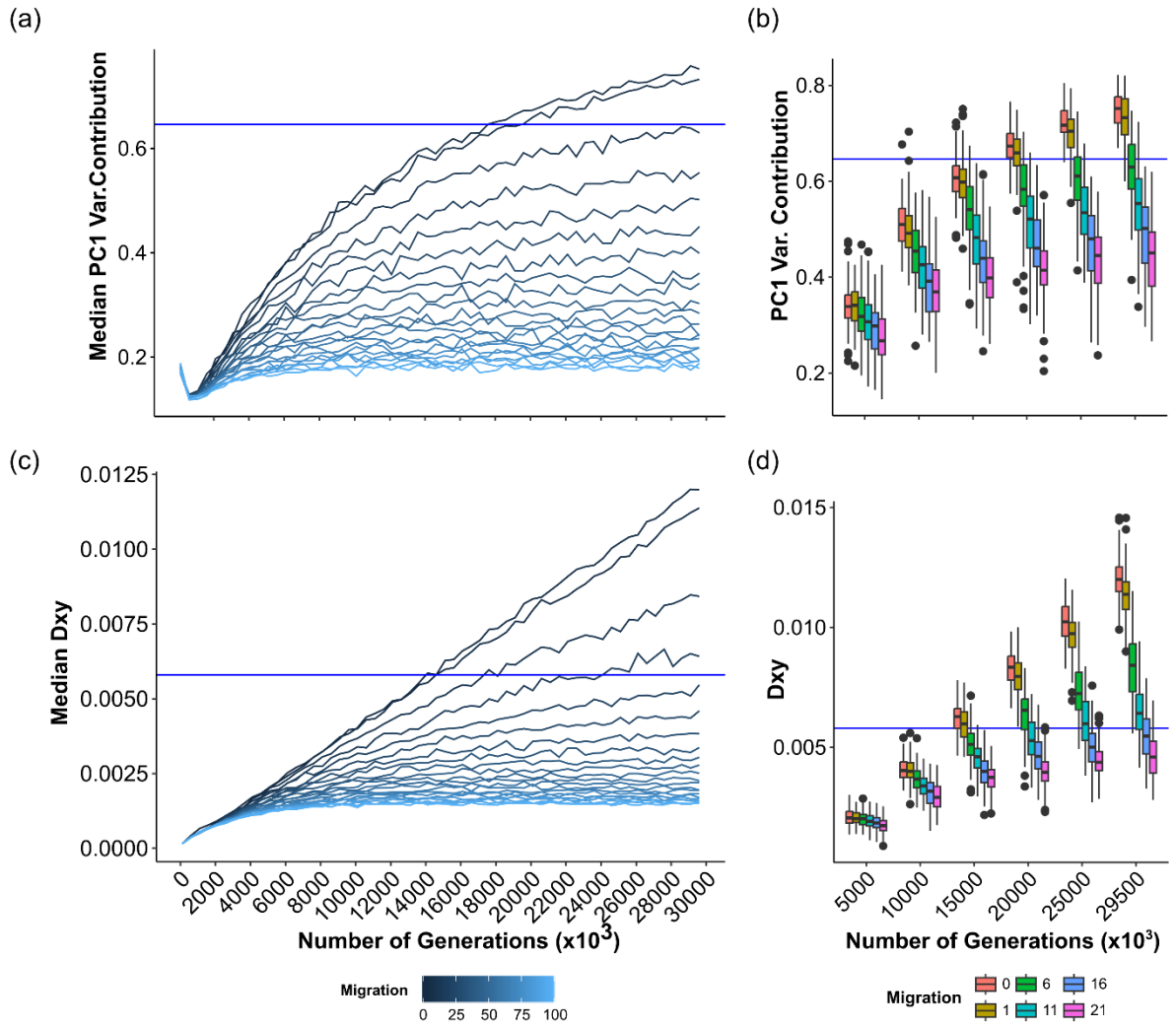
Supplementary Figure 4: Heterozygosity along the genome. Heatmap of heterozygosity per 20kb window and sample along the genome. High heterozygosity values are shown in red. Strains JB1207 and JB1169 show high heterozygosity along large parts of the genome.



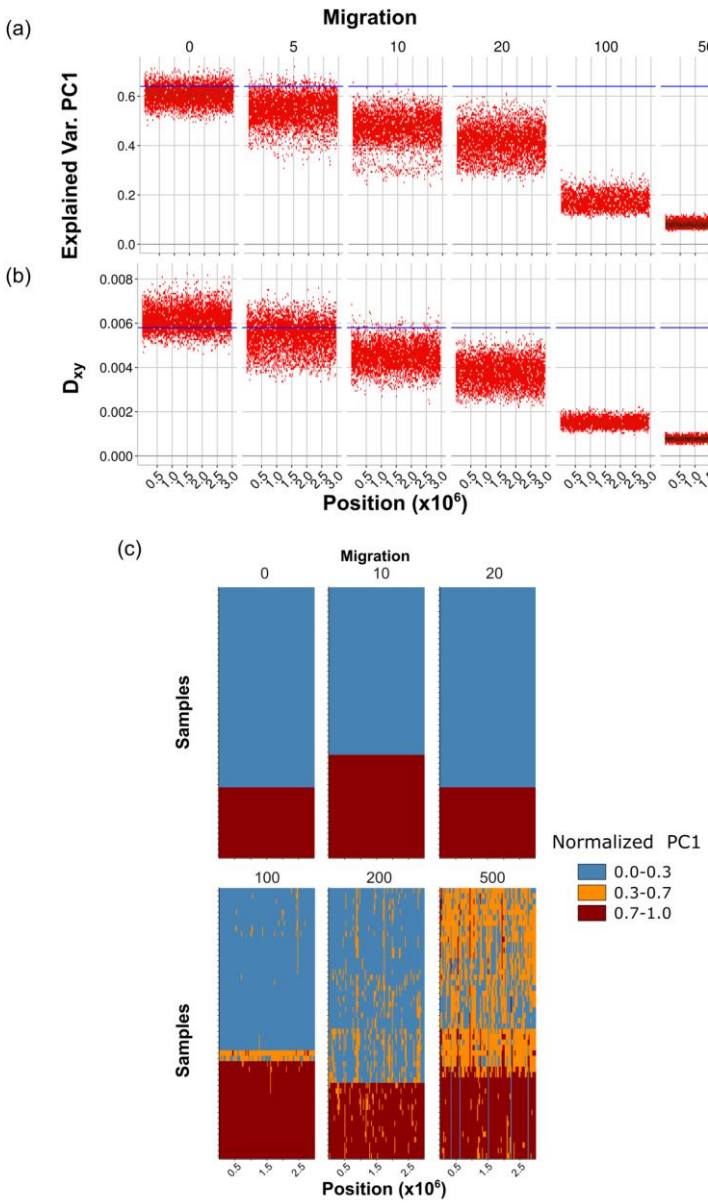
Supplementary Figure 5: Difference in genetic diversity between ancestral groups. Histogram of the ratio in genetic diversity (π) between the ancestral clade *Sk* over the *Sp* clade. Note that values are presented on a log 2 scale.


Sample
JB914
JB1199
JB763
JB1178
JB1153
JB1116
JB912
JB592
JB915
JB1203
JB872
JB858
JB852
JB846
JB841
JB839
JB857
JB842
JB851
JB918
JB1110
JB917
JB1169
JB873
JB875
JB902
JB1196
JB913
JB1180
JB838
JB943
JB884
JB862
JB1117
JB854
JB845
JB848
JB874
JB594
JB911
JB1115
JB762
JB1109
JB931
JB1174
JB942
JB930
JB934
JB932
JB933
JB948
JB910
JB1202
JB946
JB909
JB947
JB907
JB944
JB908
JB929
JB855
JB840
JB856
JB850
JB847
JB843
JB882
JB878
JB883
JB880
JB881
JB1171
JB1167
JB1154
JB1188
JB1187
JB899
JB1186
JB1185
JB893
JB593
JB4
JB905
JB859
JB1189
JB1175
JB1191
JB1190
JB1114
JB900
JB901
JB1194
JB1192
JB1193
JB871
JB837
JB866
JB758
JB916
JB1197
JB953
JB952
JB1206
JB1205
JB1207
JB853
JB849
JB844
JB1172
JB1166
JB939
JB877
JB1176
JB1108
JB1112
JB1201
JB1113
JB588
JB861
JB1198
JB876
JB1177
JB894
JB864
JB890
JB870
JB1111
JB898
JB879
JB1179
JB1168
JB1204
JB868
JB892
JB937
JB761
JB936
JB50
JB374
JB1170
JB941
JB906
JB945
JB22
JB940
JB887
JB760
JB891
JB888
JB886
JB885
JB889
JB938
JB869
JB1182
JB1184
JB759
JB897
JB1181
JB1195
JB1183
5.0e+05
1.0e+06
1.5e+06
2.0e+06
2.5e+06
3.0e+06
3.5e+06
4.0e+06
4.5e+06
5.0e+06
5.5e+06
5.0e+05
1.0e+06
1.5e+06
2.0e+06
2.5e+06
3.0e+06
3.5e+06
4.0e+06
4.5e+06
5.0e+05
1.0e+06
1.5e+06
2.0e+06

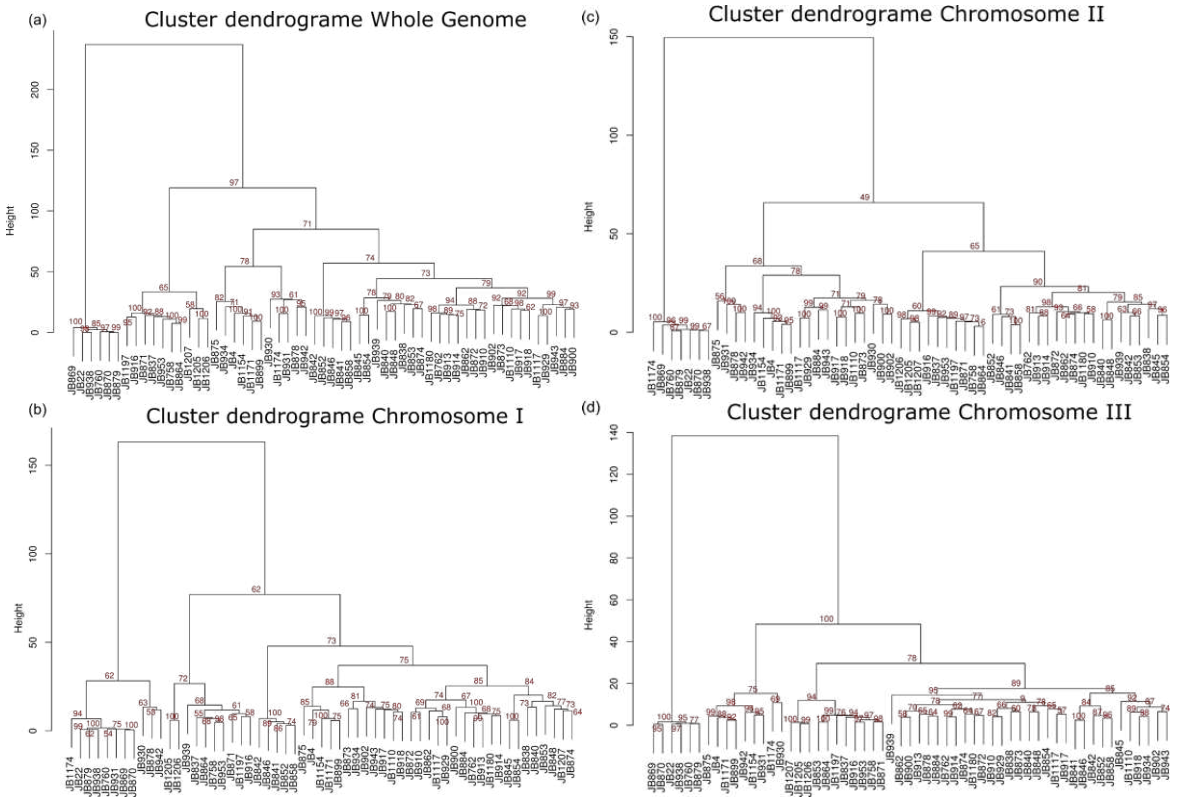
Supplementary Figure 6: Distribution of ancestral *Sp* and *Sk* haplotypes along the genome. Heatmap of ancestral *Sp* and *Sk* haplotypes along the genome for each sample. This figure corresponds to Figure 1c of the main manuscript, but uses all 161 samples.



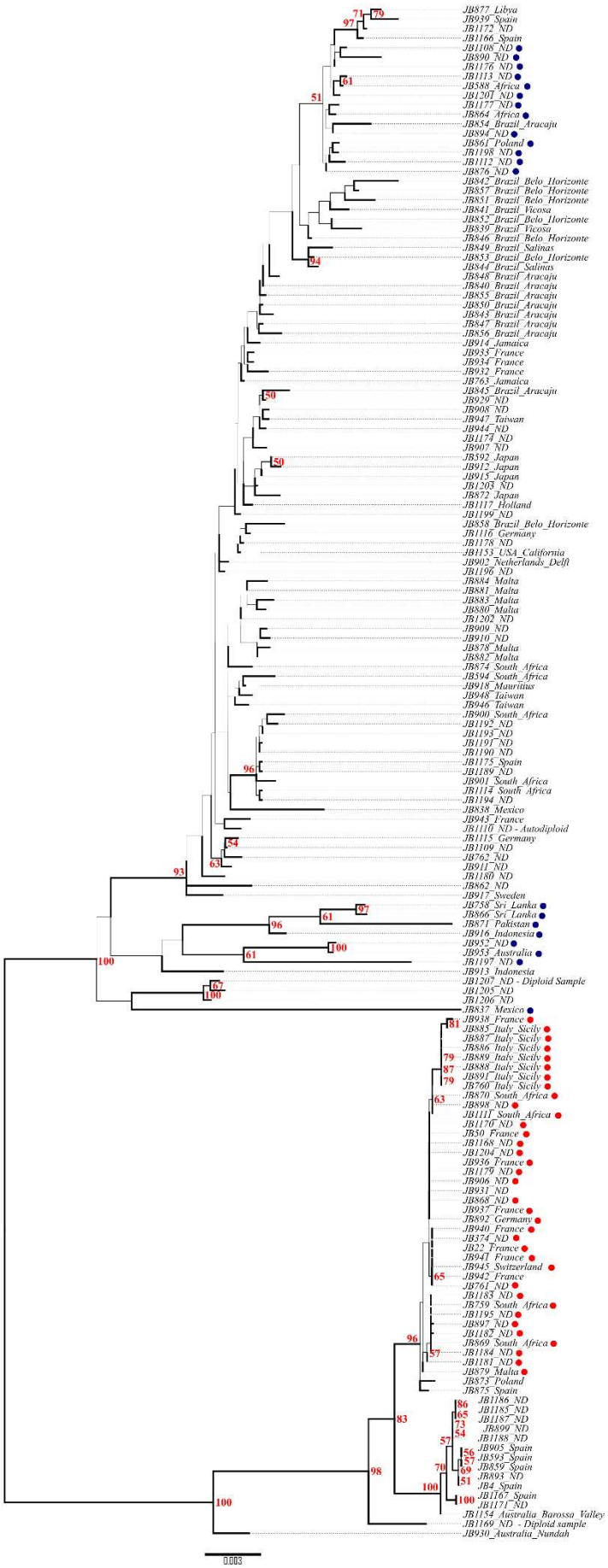
Supplementary Figure 7: Individual based forward simulations of population divergence for a single genomic window. Proportion of variance explained by PC1 over time (generations) (a-b) and divergence between subpopulations (Dxy in c-d), with different migration rates (in migrants per generation in a population of $1 \cdot 10^6$). In all panels, the horizontal blue line corresponds to the observed empirical values. In a) and c) each line corresponds to the mean value from 100 simulations. Panels b) and d) are boxplots of all 100 simulations for migration rates below 20. These plots show that observed values are only achieved under migration rates below ~20 migrants per generation.



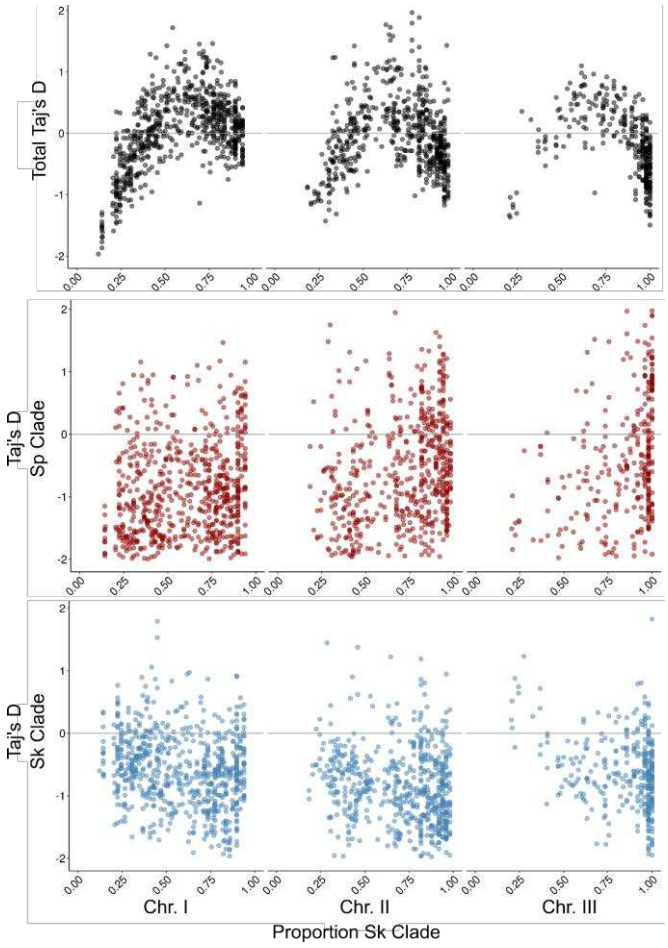
Supplementary Figure 8: Individual base forward simulations for an entire chromosome with different levels of migration. (a) Proportion of variance explained by PC1 and (b) divergence between subpopulations along simulated chromosomes of 3 Mb, after $15 \cdot 10^6$ generations for 100 independent simulations. Different migration rates were used (in migrants per generation in a population of $1 \cdot 10^6$). Horizontal blue lines correspond to observed empirical values. (c) A representative example simulation of the resulting genomes, divided into genomic windows and processed as empirical data to obtain a matrix as in Figure 1C. These plots show that the observed empirical values are only achieved under low migration rates (below ~20 migrants per generation in (a) and (b)), but higher migration rate is needed to obtain large haplotype blocks with recombined ancestral haplotypes as seen in the empirical data. Contrary to empirical data, the observed blocks under higher migration show low divergence (Dxy below 0.002) and no clear clustering of samples between subpopulations (high proportion of windows per sample with normalized PC1 between 0.3 and 0.7: yellow regions, contrary to observations in Supplementary Figures 1 - 3). General conclusions were consistent between replicated simulations.



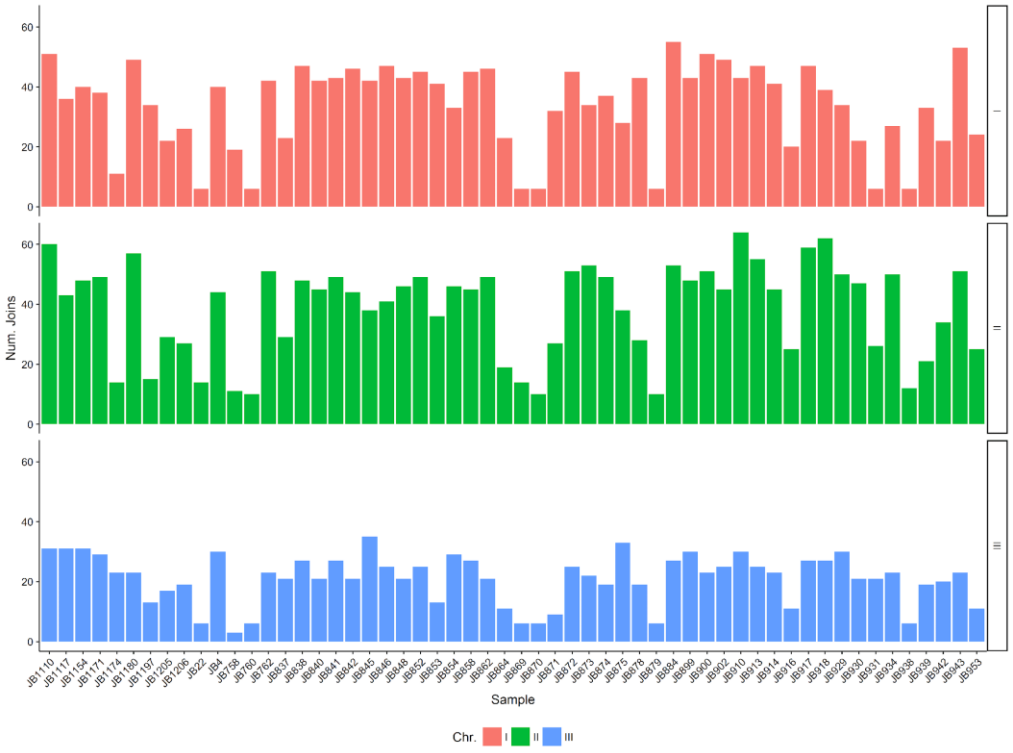
Supplementary Figure 9: Clustering analyses by chromosome. Clustering analyses using ancestral block distribution. Each panel shows the clustering using whole chromosome and independently by chromosome. Red numbers show bootstrapping support.


JB877_Libya
JB939_Spain
JB1172_ND
JB1166_Spain
JB1108_ND
JB890_ND
JB1176_ND
JB1113_ND
JB588_Africa
JB1201_ND
JB1177_ND
JB864_Africa
JB854_Brazil_Aracaju
JB894_ND
JB861_Poland
JB1198_ND
JB1112_ND
JB876_ND
JB842_Brazil_Belo_Horizonte
JB857_Brazil_Belo_Horizonte
JB851_Brazil_Belo_Horizonte
JB841_Brazil_Vicosa
JB852_Brazil_Belo_Horizonte
JB839_Brazil_Vicosa
JB846_Brazil_Belo_Horizonte
JB849_Brazil_Salinas
JB853_Brazil_Belo_Horizonte
JB844_Brazil_Salinas
JB848_Brazil_Aracaju
JB840_Brazil_Aracaju
JB855_Brazil_Aracaju
JB850_Brazil_Aracaju
JB843_Brazil_Aracaju
JB847_Brazil_Aracaju
JB856_Brazil_Aracaju
JB914_Jamaica
JB935_France
JB934_France
JB932_France
JB763_Jamaica
JB845_Brazil_Aracaju
JB929_ND
JB908_ND
JB947_Taiwan
JB944_ND
JB1174_ND
JB907_ND
JB592_Japan
JB912_Japan
JB915_Japan
JB1203_ND
JB872_Japan
JB1117_Holland
JB1199_ND
JB858_Brazil_Belo_Horizonte
JB1116_Germany
JB1178_ND
JB1153_USA_California
JB902_Netherlands_Delft
JB1196_ND
JB884_Malta
JB881_Malta
JB883_Malta
JB880_Malta
JB1202_ND
JB909_ND
JB910_ND
JB878_Malta
JB882_Malta
JB874_South_Africa
JB594_South_Africa
JB918_Mauritius
JB948_Taiwan
JB946_Taiwan
JB900_South_Africa
JB1192_ND
JB1193_ND
JB1191_ND
JB1190_ND
JB1175_Spain
JB1189_ND
JB901_South_Africa
JB1114_South_Africa
JB1194_ND
JB838_Mexico
JB943_France
JB1110_ND - Autodiploid
JB1115_Germany
JB1109_ND
JB762_ND
JB911_ND
JB1180_ND
JB862_ND
JB917_Sweden
JB758_Sri_Lanka
JB866_Sri_Lanka
JB871_Pakistan
JB916_Indonesia
JB952_ND
JB953_Australia
JB1197_ND
JB913_Indonesia
JB1207_ND - Diploid Sample
JB1205_ND
JB1206_ND
JB837_Mexico
JB938_France
JB885_Italy_Sicily
JB887_Italy_Sicily
JB886_Italy_Sicily
JB889_Italy_Sicily
JB888_Italy_Sicily
JB891_Italy_Sicily
JB760_Italy_Sicily
JB870_South_Africa
JB898_ND
JB1111_South_Africa
JB1170_ND
JB50_France
JB1188_ND
JB1204_ND
JB936_France
JB1179_ND
JB906_ND
JB931_ND
JB868_ND
JB937_France
JB892_Germany
JB940_France
JB874_ND
JB22_France
JB941_France
JB945_Switzerland
JB942_France
JB761_ND
JB1183_ND
JB759_South_Africa
JB1195_ND
JB897_ND
JB1182_ND
JB869_South_Africa
JB1184_ND
JB1181_ND
JB879_Malta
JB875_Poland
JB875_Spain
JB1186_ND
JB1185_ND
JB1187_ND
JB899_ND
JB1188_ND
JB905_Spain
JB593_Spain
JB859_Spain
JB893_ND
JB4_Spain
JB1167_Spain
JB1171_ND
JB1154_Australia_Barossa_Valley
JB1160_ND - Diploid sample
JB930_Australia_Nundah

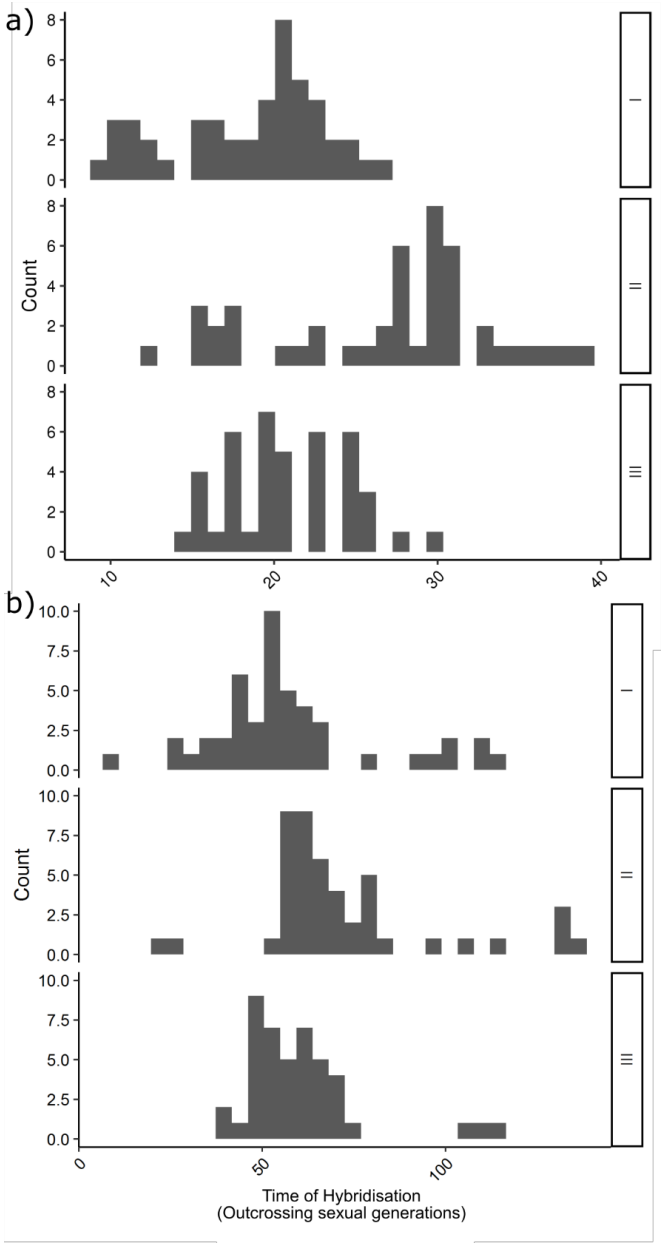
Supplementary Figure 10: Mitochondrial phylogeny. Unrooted maximum-likelihood tree for 161 samples based on mitochondrial SNPs. Red numbers show bootstrapping support when higher than 50. Tips show sample ID and sampling location as in Supplementary Table 1. Scale bar units in substitutions per site. Samples with high proportion from one ancestral population are indicated with dots (red: high *Sp* clade proportion, blue: high *Sk* clade proportion).



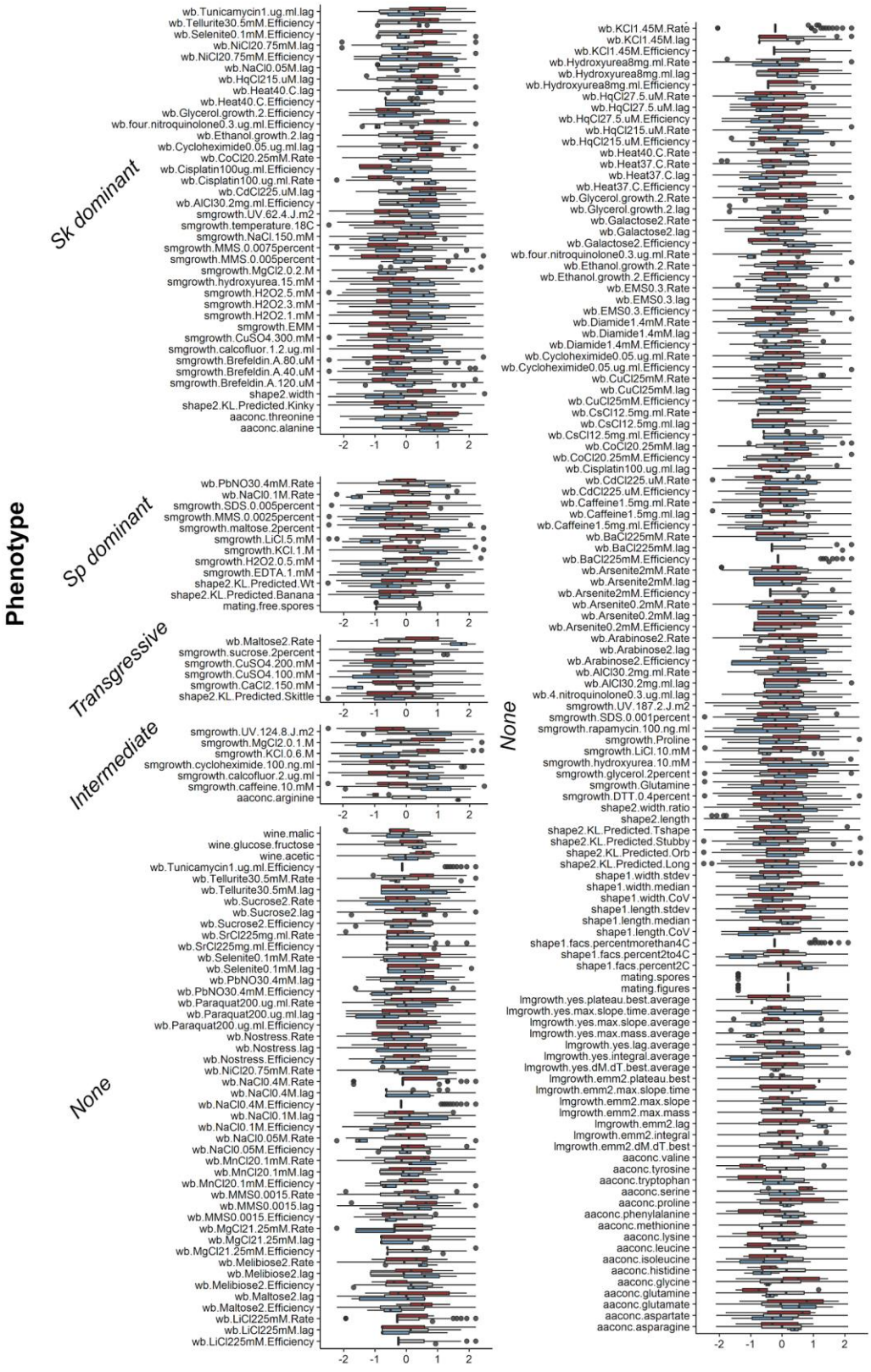
Supplementary Figure 11: Tajima's D estimate. Relationship between proportion of ancestral haplotypes (here represented by the proportion of *Sk* clade) and Tajima's D. Different panels show total Tajima's D using all samples (black dots) and samples from the *Sp* (red) or *Sk* (blue) clade only.



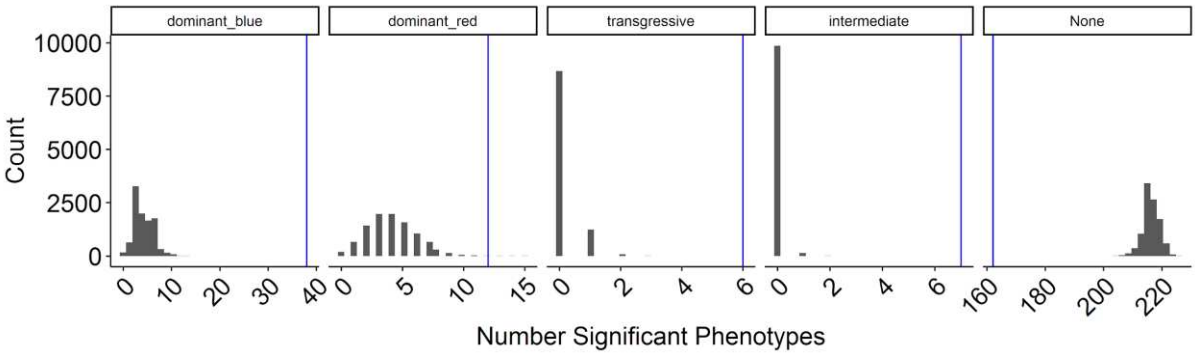
Supplementary Figure 12: Number of transitions (joins) between *Sp* and *Sk* haplotypes along the genome per chromosome and sample.



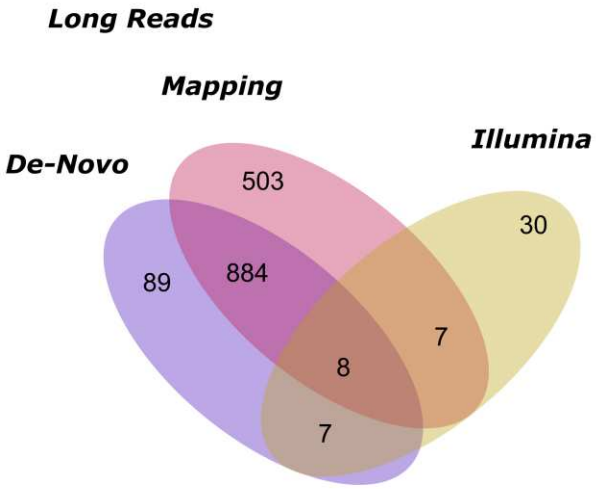
Supplementary Figure 13: Estimated number of sexual generations since hybridisation. Histogram displaying the number of outcrossing sexual generations since hybridisation using the model by Janzen *et al*. (2018). The analysis was divided by chromosome and sample. Counts represent the number of samples for which a given time of hybridization was inferred. The analysis was conducted using 200 and 50 SNPs per windows ((a) and (b) respectively).



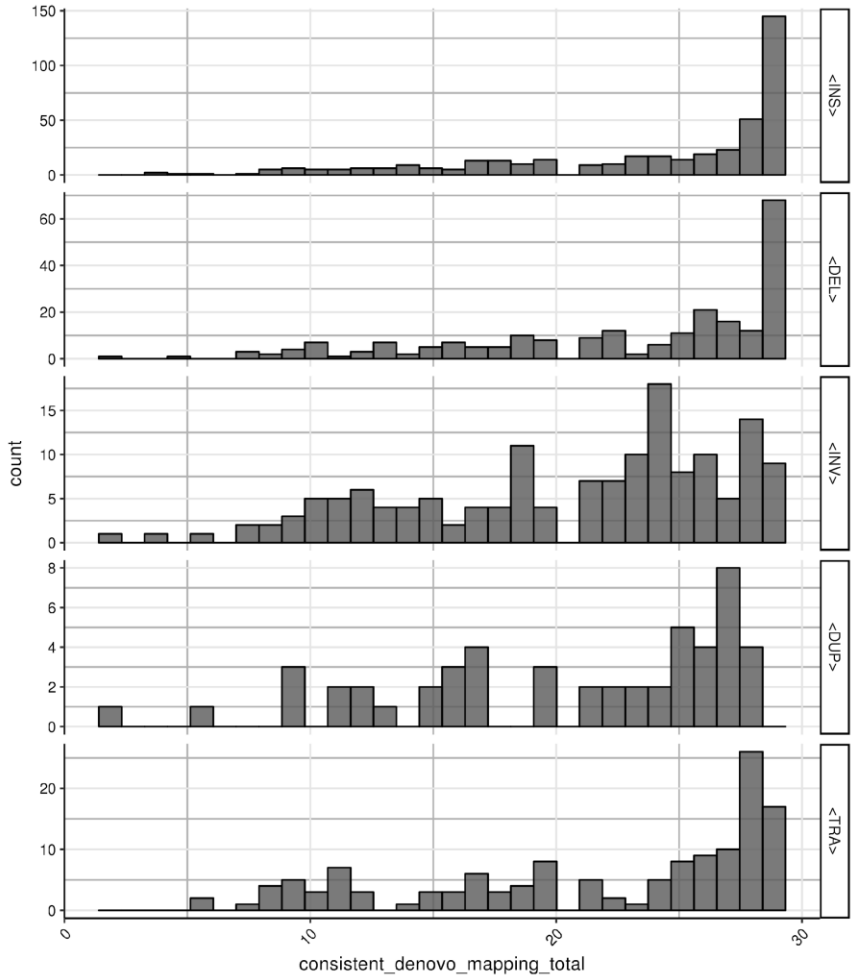
Supplementary Figure 14: Phenotypic distribution per trait. Boxplot of phenotypic distribution using normalised values. Traits are divided by ancestral admixture proportions: pure *Sk* clade (blue), pure *Sp* clade (red) and hybrids (gray).



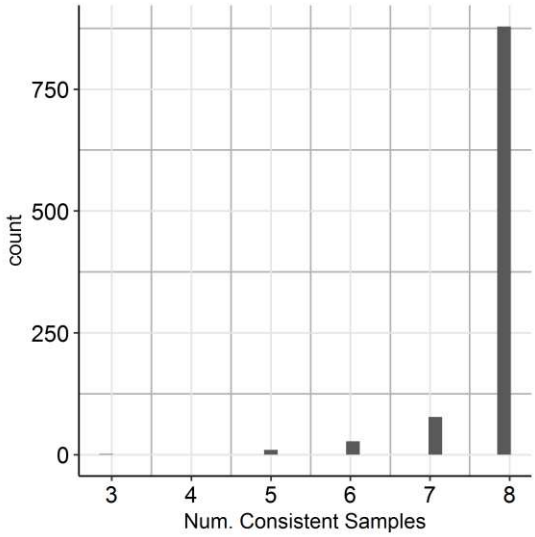
Supplementary Figure 15: Permutation test with phenotypic data. Histogram of total number of significant traits after randomising sample category (pure *Sp* clade, pure *Sk* clade and hybrids) without replacement. Blue lines show the observed number of significant traits. Matrix data was randomised 10000 times to produce the distribution per category group.



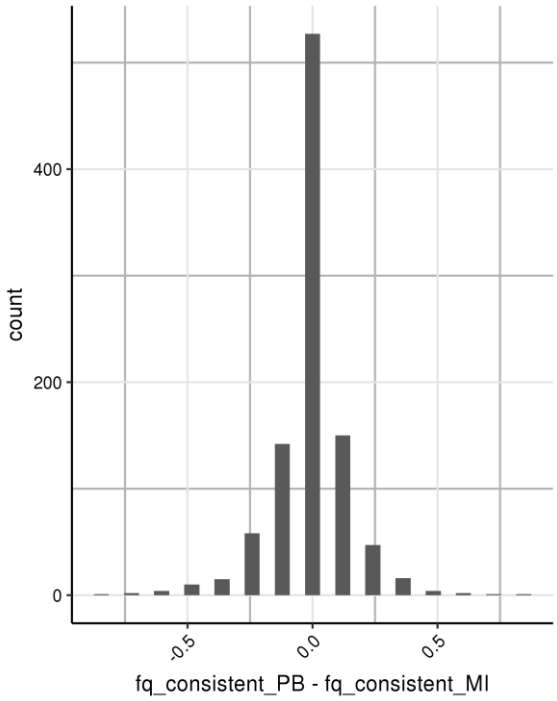
Supplementary Figure 16: Total number of structural variant calls. Venn diagram comparing the total number of structural variants in the reported list from Illumina reads (Jeffares, 2017), and long reads (this study). The latter is futher subset by the two genotype calling approaches: *de-novo* and mapping reads.



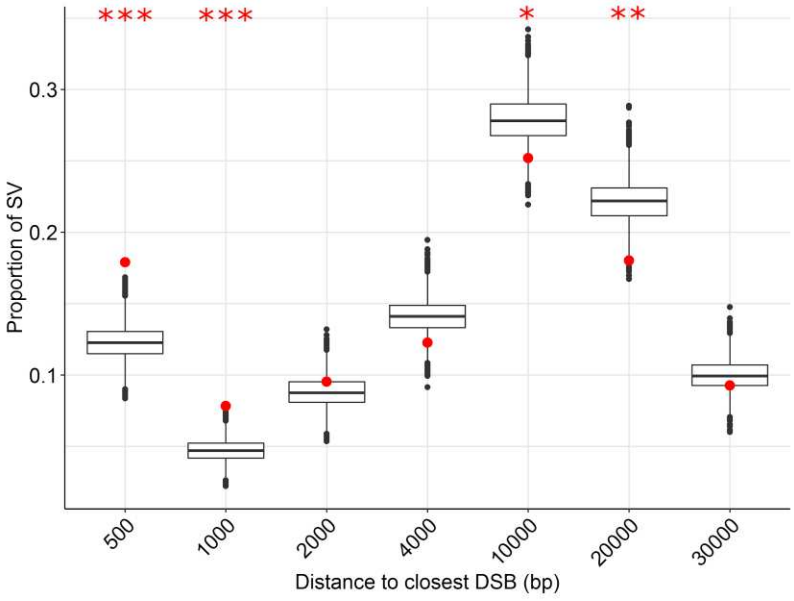
Supplementary Figure 17: Comparison between SV detection using the de-novo vs. mapping approach. Histogram of consistent genotypes of SVs in samples sequenced with PacBio and/or Nanopore (29 in total). Panels represent different types of SV (INS: insertion, DEL: deletion, INV: inversion, DUP: duplication, TRA: translocation). The obtained genotype from the de-novo and mapping approach in each sample and variant was compared. The histogram shows the number of variants with consistent mapping and de-novo genotype (no necessarily the same genotype between samples). For example, there are around 150 insertions with consistent mapping and de-novo genotype in all 29 samples. In another 50 insertion one sample out of the total shows different genotype between the mapping and de-novo approach.



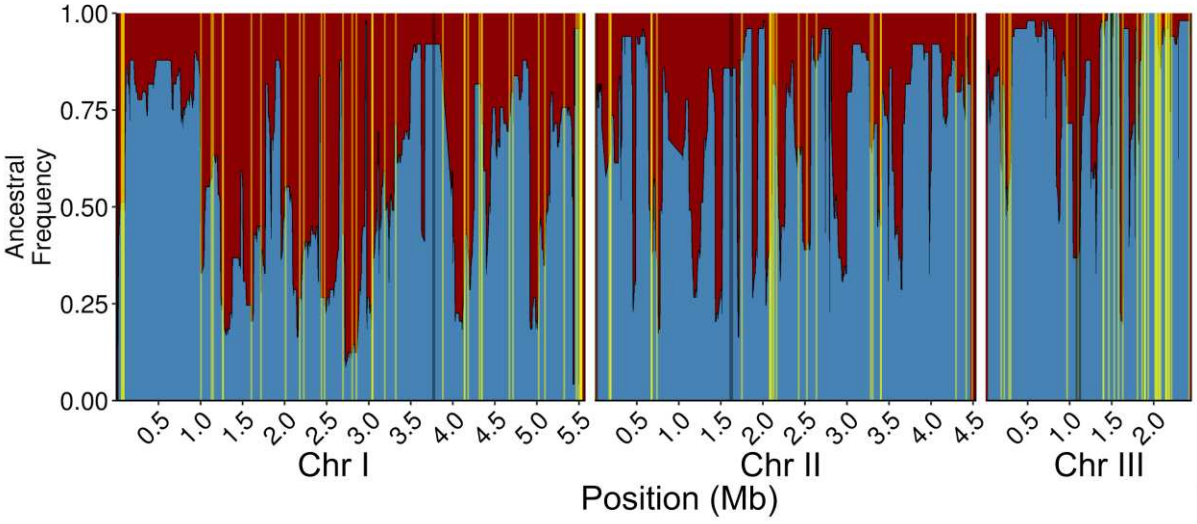
Supplementary Figure 18: Consistency of SV genotype detection using PacBio vs. MinIon data. Histogram of consistent genotypes of SVs in samples sequenced with both PacBio and Nanopore (8 strains in total). For each variants the genotype between the PacBio and Nanopore data was compared in each sample. The histogram shows the number of variants with up to 8 consistent samples. For example, 878 SVs show the same genotype per sample between PacBio and Nanopore data (not necessarily the same genotype between samples).



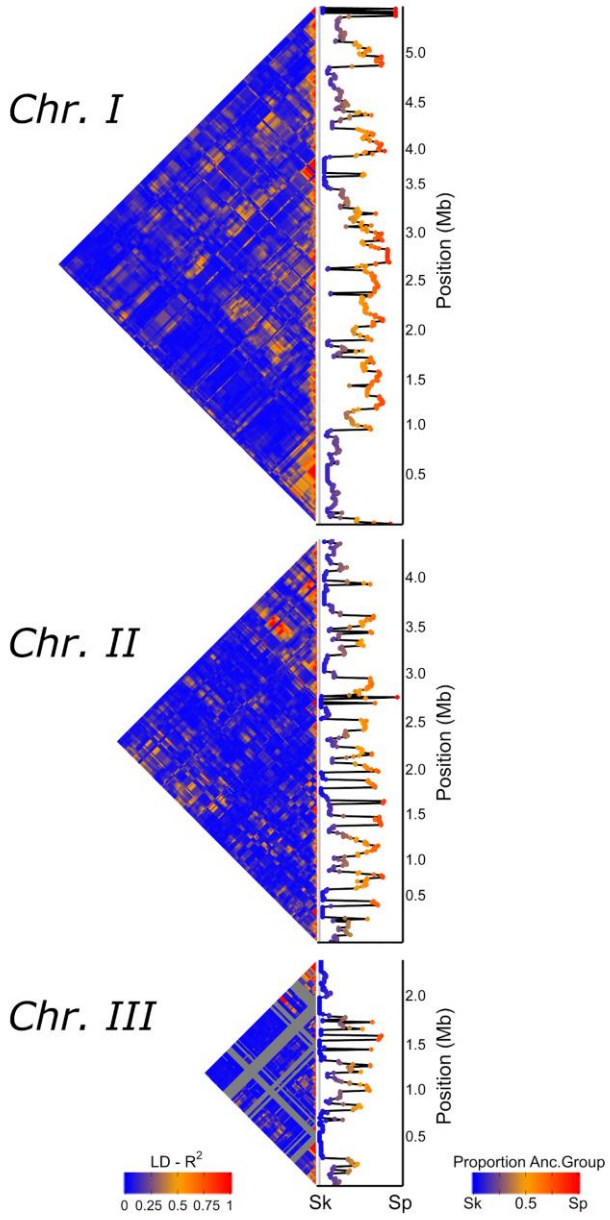
Supplementary Figure 19: Difference in allele frequency using PacBio vs. MinIon data. Histogram of the difference in observed allele frequency of samples sequenced with both PacBio and Nanopore (8 in total).



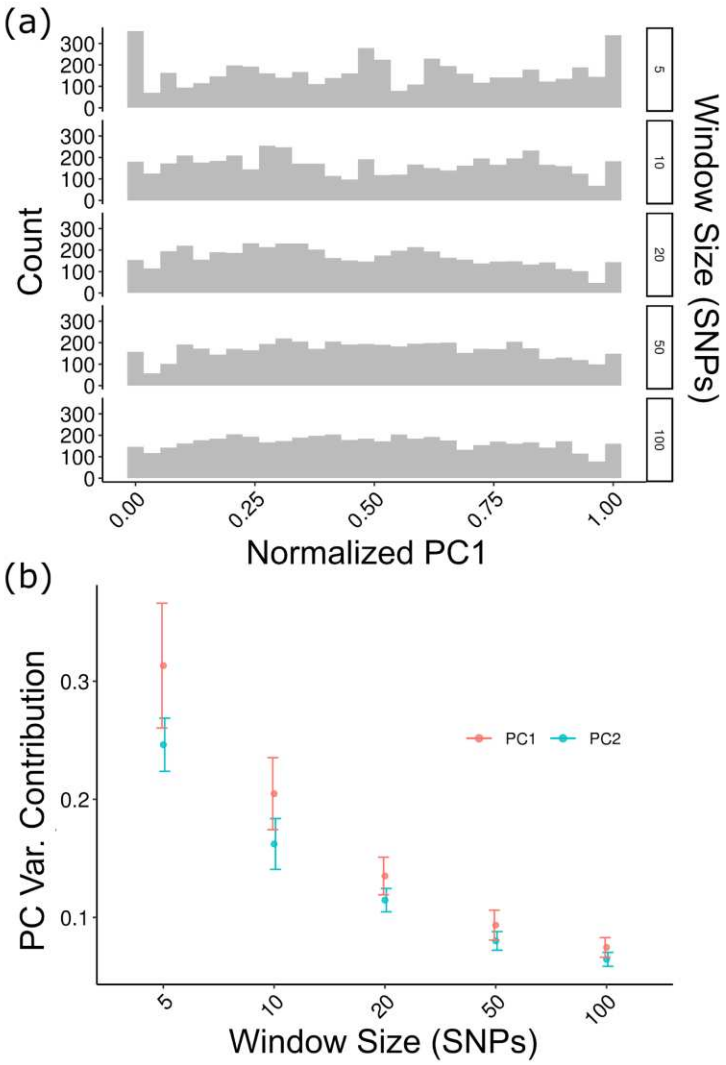
Supplementary Figure 20: Boxplot displaying the proportion of SV relative to the physical genetic distance from the closest double-strand break. Boxplot with the distribution of the expected proportion of structural variants by random permutation within different genetic distances. SV observed within 500bp and 1kb of the closes DSB are significantly overrepresented and those at 10kb or 20kb are underrepresented. Red points represent the observed proportions from long read sequencing data. Significance in difference between permutations and observed values are shown with asterisks: * p-value < 0.05, ** p-value <0.001, *** p-value <0.0001.



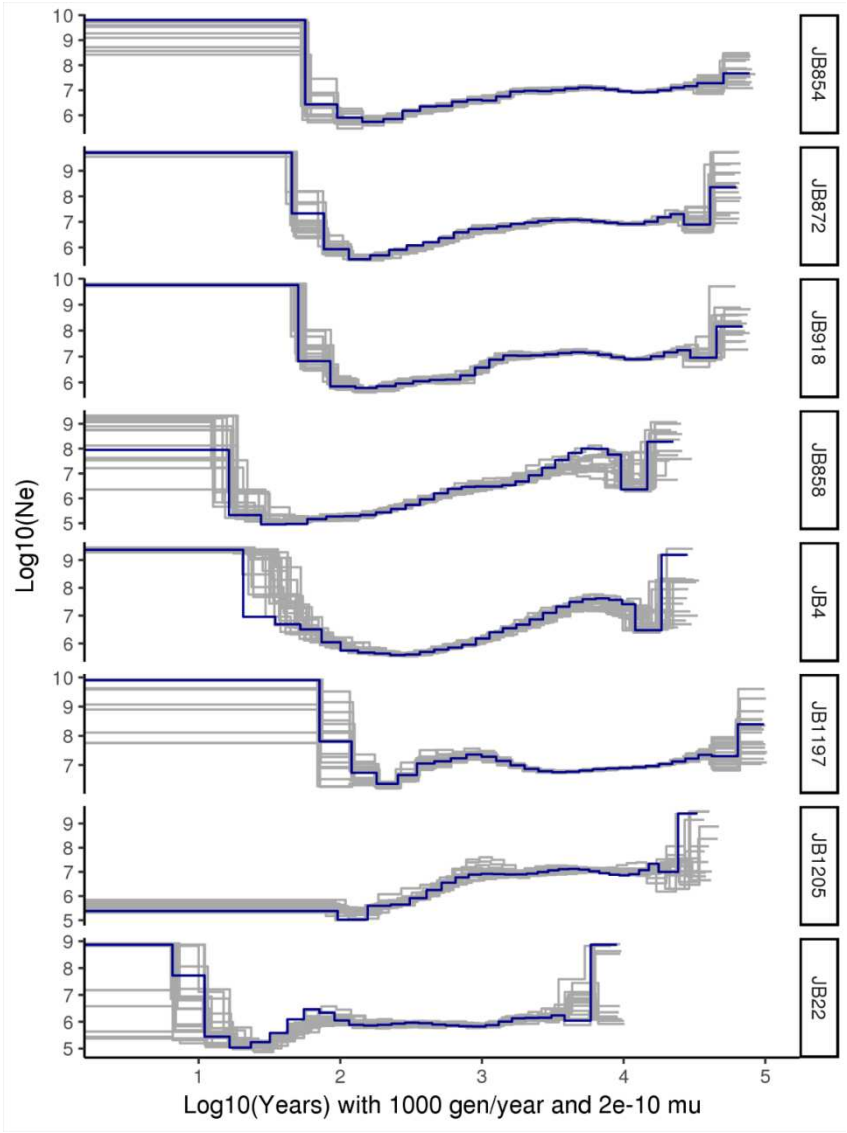
Supplementary Figure 21: Distribution of divergent SV. Proportion of *Sp* (red) and *Sk* (blue) ancestry across all 57 samples along the genome. Divergent between ancestral groups SV (with frequency difference > 0.7 between *Sp* and *Sk*) are shown with yellow lines. Centromeric regions are shown in gray.



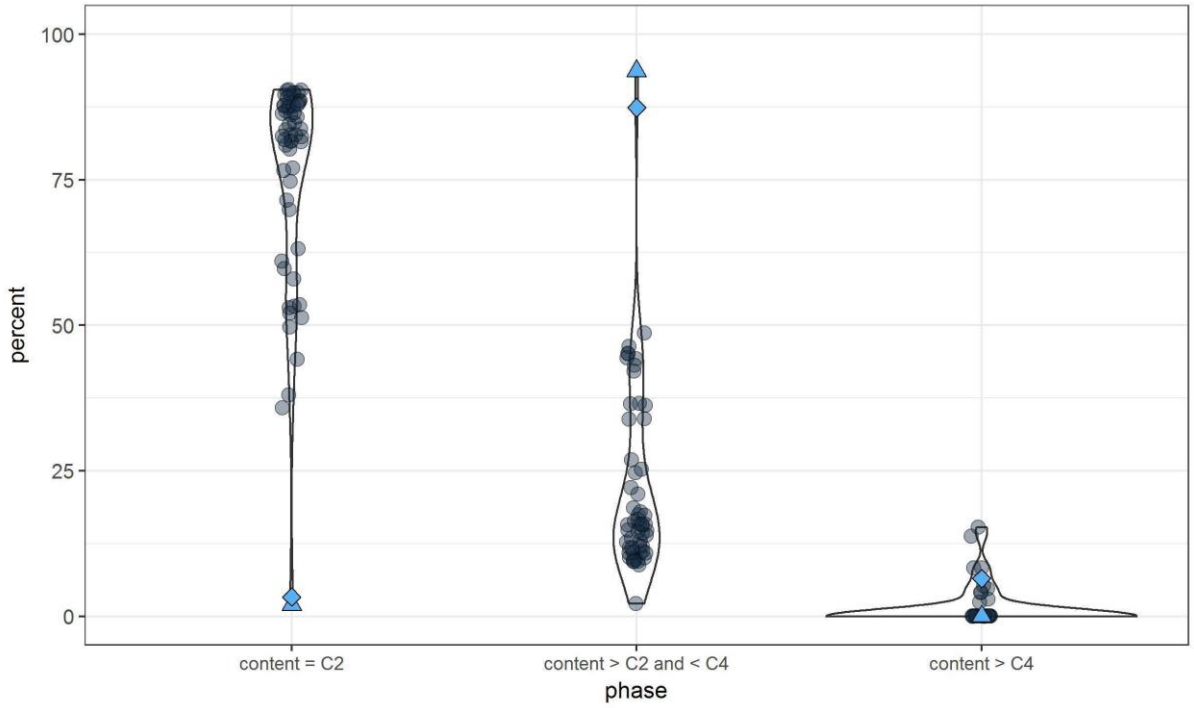
Supplementary Figure 22: Linkage disequilibrium between pairs of genomic windows polarized by ancestry. Heat map of linkage disequilibrium (R^2) for all comparison between genomic windows (left panel). Proportions of ancestral groups (*Sp* or *Sk*) along the genome are shown in the right panel. Genomic regions fixed for one of the ancestral groups are shown in gray areas in the heat map.



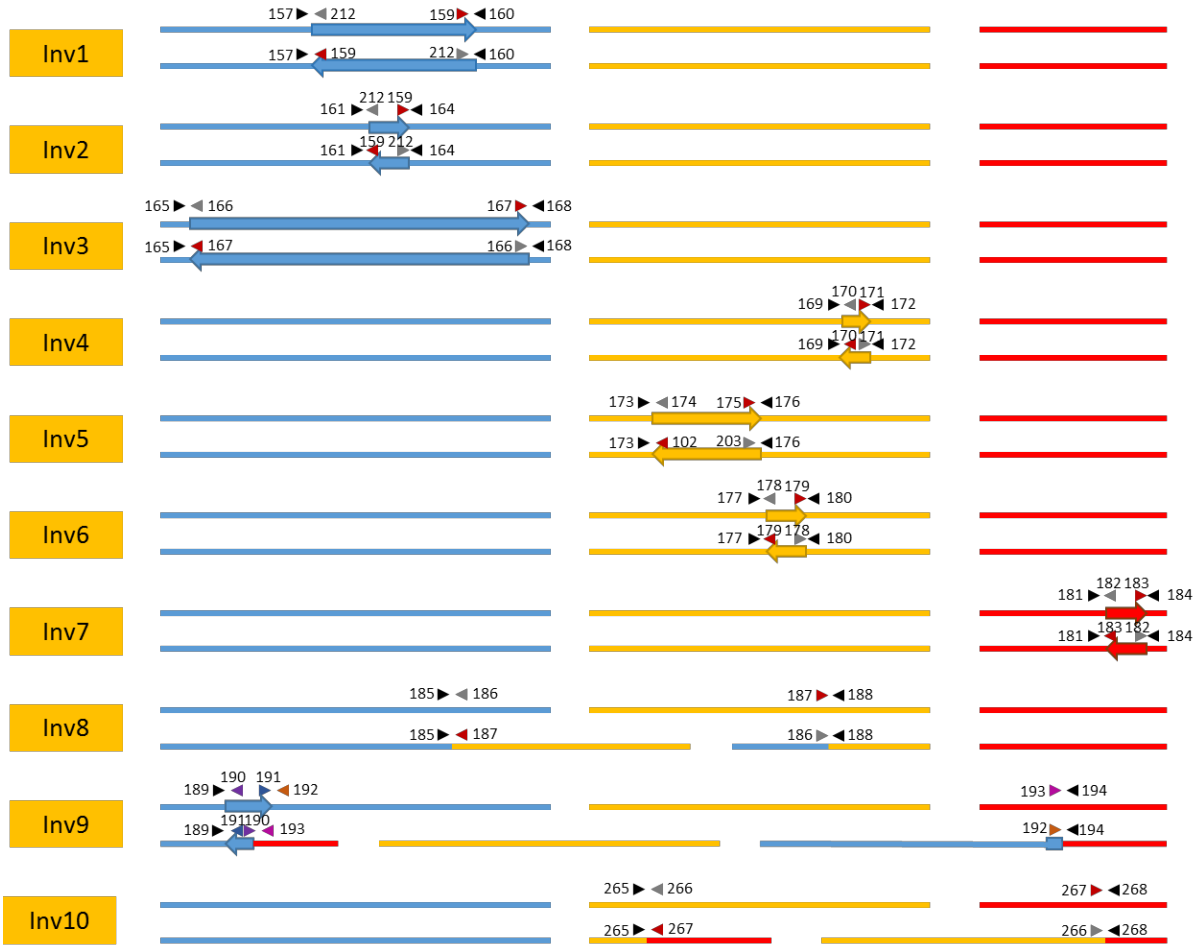
Supplementary Figure 23: Individual based forward simulations for a single genomic window of different sizes. a) Distribution of samples in the normalised PC1 for 100 replicate simulations, using window sizes from 5 to 100 SNPs. An increased number of counts for extreme values (1 and 0) are only observed in small windows (5 SNPs). This contrasts with empirical observations using window sizes of 200 SNPs where the values are almost exclusively centred around 0 and 1 (Supplementary Figure 2). b) Proportion of variance explained by PC1 and PC2 in the PCA using different window sizes. Simulations using as few as 5 SNPs per window could artificially divide a panmictic population into two clusters and explain 40 % of the variance in PC1. However, the inclusion of more SNPs per window, reduced erroneous clustering and reduced the variance to << 0.1 for 100 SNPs. None of the simulations (100 replicates per window size) approached the empirical value of 64 % of the variance explained by PC1 using a window size of 200 SNPs providing evidence that bimodal group structure was not an artefact of small window size.



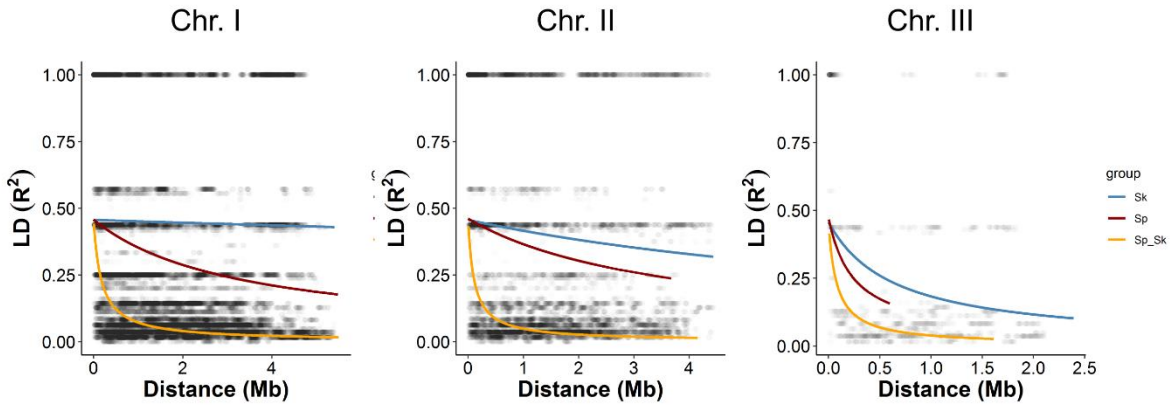
Supplementary Figure 24: Inferred effective population size over time with multiple sequentially Markovian coalescence analyses (MSMC). Analysis was divided by secondary clusters presented in Figure 1c. Each cluster is label with one representative of the cluster. Inferred values are shown with the blue line and bootstraps with grey lines. Time in x axis in log10 scale using 1000 generations per year and mutation rate of $2x10^{-10}$. Present time is in x=0 and increase in the past. Effective population size in log10 scale.



Supplementary figure 25: DNA content per strain. Representation of flow cytometry measurements of percentage of cells with nuclear DNA content <C2 (two times haploid genomic content; left), nuclear content between C2 and C4 (middle) and larger than C4 (right). Due to the short G1 phase and late cytokinesis, few cells with C1 content are found in fission yeast. Haploid cells are thus mostly found in the C2 phase. Diploid cells will be mostly in C4 or larger. The strain with high heterozygosity (JB1207) is indicated by the blue diamond and it has a large proportion of cells with large genomic content. Additionally, strain JB1169 (blue triangle) appears to be of higher ploidy, though it does not show increased heterozygosity, which suggests this strain was formed by auto-diploidization. Data from Jeffares et al. 2015.



Supplementary Figure 26: Overview of PCR primers used to verify inversions and translocations. For each SV first the positions in the reference genome are given (top), and below the organization in the rearranged form as observed in the long-read *de novo* assemblies. The three lines represent Chromosomes I, II and III. The small arrows indicate the position and direction of the primers with the numbers of the primers indicated. The large arrows indicate the size and orientation of inversions. Note that more than one SV can occur per strain, and that the representations are thus not always representations of actual strains. Primers in 'Inv10' from Zanders et al. 2014.



Supplementary figure 27: Example of decay in linkage disequilibrium (LD) with genetic distance using one representative per cluster. Relationship between LD (R2) and physical distance is depicted for each chromosome. Black points represent values for each window pair comparison. Lines show non-lineal regression model based on Hill & Weir (1988) and Remington et al. (2001). LD estimates were divided into three categories representing comparison between windows of shared ancestry (Sp-Sp or Sk-Sk) reflecting positive ancestry disequilibrium (AD) or of opposite ancestry (Sk-Sp) reflecting negative AD.